See 1473

DEPARTMENT OF STATISTICS

TECHNICAL REPORT No. 448

January, 1976

TOPICS IN TIME SERIES ANALYSIS
III. ARIMA time series models with non-Normal shocks

by

Johannes Ledolter

and

George E. P. Box

UNIVERSITY OF WISCONSIN-MADISON

Best
Available
Copy

III. ARIMA time series models with non-Normal shocks

3.1 Introduction

In this chapter we investigate the sensitivity of certain procedures used in time series analysis to non-Normality of the distribution of the shocks a_t. We suppose that the distribution function of the shocks is a member of the symmetric exponential power family, which includes the Normal as well as leptokurtic and platikurtic distributions. In particular we investigate the effect of non Normal shocks on the predictive distribution of future observations.

Wold's decomposition theorem states that every weakly stationary stochastic process can be decomposed into orthogonal shocks. The Gaussian hypothesis assumes that the shocks follow a Normal distribution with fixed mean and fixed variance. In this case the process is characterized by the first and second order moments and various estimation procedures based on the Normality assumption of the shocks have been developed.

The Normality assumption appears to be reasonable for many kinds of series. However, it was pointed out by Kendall (1953), Mandelbrot (1961, 1963, 1967), Fama (1965) that particularly for stock price data the distribution of the error terms seems to be leptokurtic.

The following study was inspired by the continuing dispute among some economists about the underlying stochastic

process for stock price series. In the next section some of the ideas are reviewed.

3.2. Statistical models for stock price series

Some speculators purport to be able to infer from charts certain technical patterns that enable profitable prediction of future price changes of stocks. "Chartist technicians however are in low repute, because they usually have holes in their shoes and no favorable records of reproducible worth" (Samuelson (1971)). Economic fundamentalists have been equally unsuccessful in predicting the future price of commodities from predetermined economic variables.

An early important contribution to the theory of stock prices was made by Louis Bachelier, who in 1900 wrote a thesis on the Theory of Speculation. His work does not only deserve an honored place in the field of economics, it also represents a major pioneering work in the area of stochastic processes (Brownian motion).

Gaussian random walk hypothesis:

Bachelier considers a Markov process in continuous time with continuous state space and with

$$P\{Z_t \le z_t | Z_0 = z_0\} = F(z_t - z_0; t) \qquad (3.2.1)$$

as a possible model for stock price series. He derives, what later became known under the name of Chapman-Kolmogorov equation,

$$F(z;t_1+t_2) = \int F(z-u;t_1)dF(u;t_2) \tag{3.2.2}$$

and shows that the Normal distribution

$$F(z;t) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{t}} e^{-\frac{1}{2\sigma^2 t}(z-\mu)^2} \tag{3.2.3}$$

satisfies this equation. Thus, he suggests the Wiener process as a possible stochastic model for stock prices series. In the discrete time analog to the Wiener process, this would mean that stock prices follow a random walk with Normal distributed errors (Gaussian random walk hypothesis).

Stable Paretian random walk hypothesis:

However the Normal distribution is not the only distribution which satisfies equation (3.2.2). It can be shown that any member of the class of stable distributions will also satisfy this equation. Stable distributions have no general explicit form for the probability density function. They can be characterized by the characteristic function which is given by

$$\phi(u) = \int_{-\infty}^{+\infty} e^{iuz} dF(z) = \begin{cases} \exp\{i\delta u - \gamma|u|^{\alpha}[1 + \frac{i\xi u}{|u|}\tan(\frac{\alpha\pi}{2})]\} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi}\log|u| & \text{if } \alpha = 1 \end{cases} \tag{3.2.4}$$

where: $-\infty < \delta < \infty$ $\quad 0 < \alpha \leq 2$

$\gamma \geq 0$ $\quad -1 \leq \xi \leq 1$

δ, γ, α, ξ are the location parameter, scale parameter,

characteristic exponent and index of skewness, respectively.

For the following three special cases explicit forms for the probability density function can be given:

i) $(\alpha=2, \xi=0)$ .... Normal distribution

ii) $(\alpha=1, \xi=0)$ .... Cauchy distribution

iii) $(\alpha=1/2, \xi=1)$ .... inverted Chi square distribution

The family of stable distributions has finite moments up to order $r < \alpha$, except in the case $\alpha = 2$, where all moments are finite. If $\alpha < 2$ the distribution is leptokurtic. ξ is an index of skewness; for $\xi = 0$ the distribution is symmetric around δ.

The family of stable distributions has the following important properties:

i) Stability or invariance under addition: This means that any finite sum of independent random variables which follow a stable distribution with parameters α and ξ is again from the class of stable distribut with the same value of α and ξ.

ii) Generalized central limit theorem: This means that this family contains all the limiting distributions for sums of independent identical distributed rando variables.

The only distribution from this class with finite varia

s the Normal distribution. If $\alpha < 2$ the variance is nfinite and the sample standard deviation becomes meaningless.

Empirical studies of stock price data show that uccessive differences of the stock prices are nearly ndependent, thus confirming the random walk hypothesis. The series looks like a wandering one, almost as if once week the Demon of Chance drew a random number from a ymmetrical population of fixed dispersion and added it to he current price to determine the next week's price" Kendall (1953)). However it was also pointed out for xample by Kendall (1953) and Mandelbrot (1961, 1963, 1967) hat the Gaussian hypothesis is subject to considerable oubt, since the distribution of the error terms seems to be eptokurtic. This and the theoretical reasons mentioned bove led Mandelbrot to adopt the stable Paretian random alk hypothesis which he developed in a sequence of papers. e assumed that differences of stock prices follow a stable istribution with characteristic exponent α in the interval $\leq \alpha < 2$.

The stable Paretian hypothesis has important implications or data analysis. Whenever $\alpha < 2$, the variance is infinite, nd the sample standard deviation which is used to measure isk becomes meaningless. Furthermore other statistical ools which are based on the assumption of finite variance re considerably weakened. Some doubt has been thrown on

this hypothesis by other authors. For example Hsu, Miller and Wichern (1974) point out that the stable Paretian random walk hypothesis does not agree with results from many stock price series observed in practice.

It is unrealistic to expect that variation in stock prices would be independent of the level. It is obviously much more natural to expect the variation to be approximately constant after the log transformation is applied to the original series. From an empirical standpoint it is reassuring to know that, if one estimates the transformation from the class of power transformations $T(z_t) = z_t^{\lambda}$, as considered by Box and Cox (1964), the estimated transformation will be very close to the log transformation. This was checked by the author for the stock price series listed in Hsu (1973).

Other possible models for stock price series:

It is not necessary to assume the Paretian hypothesis to explain the leptokurtic shock distributions found in practice. In the literature on economic stock price series various other characterizations have been put forward.

Press (1967) considered a mixture of Normal distributions with different variances to explain the heavy tailed distribution of the errors. Praetz (1972) suggested a scaled t distribution to achieve leptokurtic error distributions. Miller, Wichern and Hsu (1972) give another

interesting interpretation for heavy tailed error distributions. Instead of characterizing the errors by leptokurtic distributions, they relax the stationarity assumption of the model. They investigate long records of stock price data and mere plot of $w_t = (1-B)\log z_t$ indicates the presence of possible changes in the parameters of the model over time. These changes can be either changes in the variance of the shocks or changes in the parameter θ of the integrated moving average model $w_t = (1-B)\log z_t = (1-\theta B)a_t$.

Parameter changes can be thought of as step changes, where the parameter shifts from one level to the other at some time point t, or as occuring at every time point t, thus assuming a stochastic model for the parameters. In Appendix 3.1 this aspect is investigated further. There we consider the following three models:

Model 1: $w_t = (1-\theta B)a_t$ $\quad Ea_t = 0$ for all t

$$Ea_t^2 = \begin{cases} \sigma_1^2 & 1 \le t \le T_1 \\ \sigma_2^2 & T_1 + 1 \le t \le T_1 + T_2 \end{cases}$$

Model 2: $w_t = (1-\theta_t B)a_t$ $\quad Ea_t = 0$ and $Ea_t^2 = \sigma_a^2$ for all t

$$\theta_t = \begin{cases} \theta_1 & 1 \le t \le T_1 \\ \theta_2 & T_1 + 1 \le t \le T_1 + T_2 \end{cases}$$

Model 3: $w_t = (1-\theta_t B)a_t$

$\phi(B)(\theta_t - \theta) = \psi(B)\alpha_t$ $\quad Ea_t = E\alpha_t = 0$

$Ea_t^2 = \sigma_a^2$; $E\alpha_t^2 = \sigma_\alpha^2$ for all t

$\{a_t\}$ and $\{\alpha_t\}$ are independent white noise sequences .

It is shown in Appendix 3.1 that in everyone of these cases the errors e_t, which are evaluated from the model $w_t = (1-\theta B)a_t$ (i.e.: assuming constant parameters) will appear to come from a leptokurtic distribution. Thus, ignoring the changes in the parameters, one would be led to consider leptokurtic error distributions.

It appears that in the case of the empirical time series considered by Hsu (1973) non stationarity of the parameters is a very reasonable way to explain the data. In this thesis, however, we consider the consequences of a different hypothesis. We assume the usual form of the ARIMA model with constant parameters, but allow the possibility that the error distribution is symmetric and possibly non Normal, but not a stable distribution. Instead we assume that it is a member of the exponential power family and thus could be leptokurtic or platikurtic. We will investigate the effect of such non Normality on

i) estimation of parameters in ARIMA models

ii) forecasting of future observations

3.3 ARIMA time series models with shocks from the family of symmetric exponential power distributions

We consider the linear filter model

$$z_t = \psi(B)a_t \tag{3.3.1}$$

where

$$\psi(B) = \frac{\theta_q(B)}{\phi_p(B)(1-B)^d}\,.$$

The shocks a_t are independent drawings from the family of symmetric exponential power distributions with probability density function

$$p(a) = \omega(\beta)\sigma^{-1}\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\,|a|^{2/1+\beta}\} \tag{3.3.2}$$

where

$$\omega(\beta) = \frac{\{\Gamma(\frac{3}{2}(1+\beta))\}^{1/2}}{(1+\beta)\{\Gamma(\frac{1}{2}(1+\beta))\}^{3/2}}$$

and

$$c(\beta) = \left\{\frac{\Gamma(\frac{3}{2}(1+\beta))}{\Gamma(\frac{1}{2}(1+\beta))}\right\}^{\frac{1}{1+\beta}}$$

σ $(\sigma>0)$ is the standard deviation of the population and β $(-1<\beta\leq 1)$ is a measure of the kurtosis indicating the extent of non Normality of the parent distribution of the shocks. If $\beta = 0$ the shocks are Normally distributed.

$\beta > 0$ will result in a leptokurtic distribution;

$$\gamma_2(a) = \frac{Ea^4}{[Ea^2]^2} - 3 > 0$$

$\beta < 0$ will give rise to a platikurtic distribution;

$$\gamma_2(a) = \frac{Ea^4}{[Ea^2]^2} - 3 < 0$$

This family of distributions ranges from the uniform distribution $(\beta \to -1)$ to the double exponential distribution $(\beta=1)$.

Box and Tiao use this family of distributions extensively. The shape of the distribution and the 100 α percent points of the exponential power distribution for various values of β in units of the standard deviation σ give further insight and are given in Box and Tiao (1973).

In time series analysis the usual assumption is that shocks come from a Normal distribution $(\beta=0)$. In this part of the thesis we characterize the error distribution through one additional parameter, β, thus broadening the model. Proceeding in this way has proved useful in studies of what has been called inference robustness. Box and Tiao (1962) distinguish between two types of robustness: criterion robustness and inference robustness. Criterion robustness is concerned with how sensitive the distributional properties of a criterion are to changes in the underlying assumptions of the model. Inference robustness is concerned with how

inferences are affected when assumptions are changed.

The symmetric exponential power distributions are of course not stable distributions and in Appendix 3.2 we give an expression for the moments and for the characteristic function.

For $\beta \neq 0$ the distribution of z_t of model (3.3.1) is complicated. However for any given stationary process the value of the kurtosis of the distribution of z_t can be calculated.

It is easily shown that the kurtosis for the stationary process $z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$ is given by

$$\gamma_2(z) \doteq \frac{E(z^4)}{[E(z^2)]^2} - 3 = \frac{\sum_{j=0}^{\infty} \psi_j^4}{[\sum_{j=0}^{\infty} \psi_j^2]^2} \gamma_2(a) \qquad (3.3.3)$$

Box (1953) showed that the kurtosis for the family of exponential power distributions is given by

$$\gamma_2(a) = \frac{\Gamma(\frac{5}{2}(1+\beta))\Gamma(\frac{1}{2}(1+\beta))}{[\Gamma(\frac{3}{2}(1+\beta))]^2} - 3 \qquad (3.3.4)$$

Applying the Schwarz inequality we see that

$$|\gamma_2(z)| \leq |\gamma_2(a)| \qquad (3.3.5)$$

Thus, $\gamma_2(z)$ is always closer to its Normal value of zero than is $\gamma_2(a)$. We later illustrate this point in more detail with examples.

3.4 Parameter estimation for autoregressive models of order p with shocks from the family of symmetric exponential power distributions

We consider the process

$$(1-\phi_1 B-\phi_2 B^2-\dots-\phi_p B^p)z_t = a_t \qquad (3.4.1)$$

where z_t is a stationary difference of original observations with $Ez_t = 0$. Furthermore we assume that the roots of $\phi(B) = 1-\phi B-\dots-\phi_p B^p = 0$ lie outside the unit circle. The a_t are assumed independent with distribution given in (3.3.2).

We derive the likelihood function for the parameters in model (3.4.1) conditional on given starting values $\underset{\sim}{z}_p' = (z_1, z_2, \dots, z_p)$

$$p(a_{p+1}, a_{p+2}, \dots, a_n | \sigma, \beta)$$

$$= [\omega(\beta)]^{(n-p)} \sigma^{-(n-p)} \exp\left(-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=p+1}^{n} |a_t|^{\frac{2}{1+\beta}}\right) \qquad (3.4.2)$$

Making the transformation

$$z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = a_t \quad \text{for } p+1 \leq t \leq n$$

and treating $\underset{\sim}{z}_p$ as given, we derive

$$p(\underset{\sim}{z}|\sigma,\beta,\underset{\sim}{\phi},\underset{\sim}{z}_p) = [\omega(\beta)]^{(n-p)}\,\sigma^{-(n-p)}$$

$$\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\sum_{t=p+1}^{n}|z_t-\phi_1 z_{t-1}-\dots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}\} \qquad (3.4.3)$$

where:

$$\underset{\sim}{z}' = (z_{p+1},z_{p+2},\dots,z_n)$$

First we derive the posterior distribution of the parameters σ and $\underset{\sim}{\phi}' = (\phi_1,\phi_2,\dots,\phi_p)$ for a specific parent distribution of the shocks, thus considering β fixed. For given β and for fixed starting values $\underset{\sim}{z}_p$, the likelihood function of $(\sigma,\underset{\sim}{\phi})$ is given by

$$\ell(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto \sigma^{-(n-p)}$$

$$\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\sum_{t=p+1}^{n}|z_t-\phi_1 z_{t-1}-\dots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}\} \qquad (3.4.4)$$

Bayes formula states that the posterior distribution of $(\sigma,\underset{\sim}{\phi})$ is proportional to

$$p(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto p(\sigma,\underset{\sim}{\phi})\,\ell(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta)$$

where $\ell(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta)$ is the likelihood function in (3.4.4) and $p(\sigma,\underset{\sim}{\phi})$ is a chosen prior distribution.

An analysis would usually be required in circumstances where little was assumed to be known about the parameters a priori. The question how a prior should be chosen so as to be "non informative" has been subject of considerable research, speculation and arguments. In particular in cases where it is applicable one can use Jeffreys' principle to derive a non informative prior distribution for the parameters According to his rule the prior distribution for the parameters $(\sigma,\underset{\sim}{\phi})$ should be chosen proportional to the square root of Fisher's information matrix.

As shown in appendix 3.3, in the present instance, provided $\beta < 0$, Jeffreys' principle leads to a prior distribution of the form

$$p(\sigma,\underset{\sim}{\phi}) = p(\sigma)p(\underset{\sim}{\phi}) \propto \sigma^{-1}\,|\Gamma_p|^{\frac{1}{2}} \qquad (3.4.5)$$

where $P_I = \{\rho_{|i-j|}\}$ is a p×p autocorrelation matrix with elements

$$\rho_{|i-j|} = \frac{E(z_{t-i}z_{t-j})}{E(z_t^2)} \text{ for } 1 \le i,\ j \le p$$

The second factor in the above expression (3.4.5) becomes important only for autoregressive parameters approaching non stationarity. Figure 3.1 shows the form of the prior distribution implied by Jeffreys' rule for a first order autoregressive process.

$$p(\sigma,\phi) = p(\sigma)p(\phi) \propto \sigma^{-1} \frac{1}{\sqrt{1-\phi^2}}$$

Over the range $-.8 \leq \phi \leq .8$ the prior $p(\phi)$ appears sufficiently flat compared with the likelihood and can be considered constant.

In the following we therefore use the simple approximation

$$p(\sigma,\underset{\sim}{\phi}) \propto \sigma^{-1} \tag{3.4.6}$$

Combining the prior distribution in (3.4.6) with the likelihood (3.4.4) we derive the posterior distribution

$$p(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto \sigma^{-(n-p+1)}\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\,S(\underset{\sim}{\phi};\beta)\} \tag{3.4.7}$$

where $S(\underset{\sim}{\phi};\beta) = \sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\dots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}$

Integration over the nuisance parameter σ in (3.4.7) gives the posterior distribution

$$p(\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta) = \{S(\phi;\beta)\}^{-\frac{n-p}{2}(1+\beta)} \tag{3.4.8}$$

From (3.4.8) it is clear that the posterior distribution of $\underset{\sim}{\phi}$ depends heavily on the value of β chosen. We shall illustrate this in some detail later. However, this does not necessarily mean that for a given body of data the inferences will be imprecise. This is because considering β now as a random variable it itself will possess a posterior distribution. It is oft.n the case

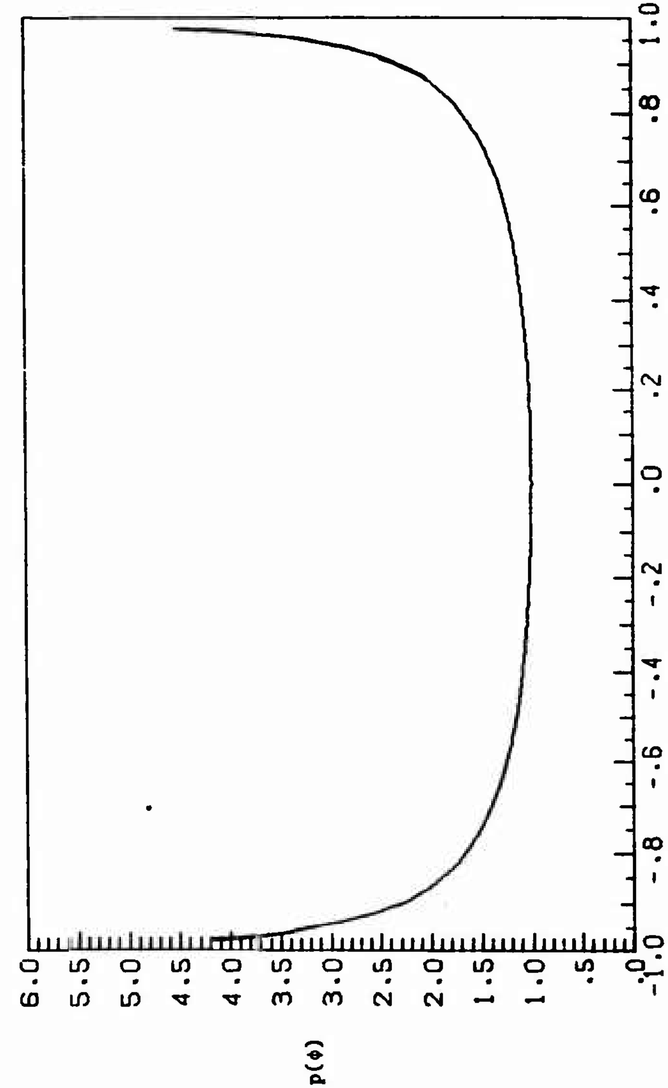


Figure 3.1: Jeffrey's non informative prior distribution for the first order autoregressive model with shocks from the exponential power

in time series work that the number of observations is rather large; in particular often more than 100 observations are available. Thus some rather precise information about β can be supplied by the data and may be incorporated in the analysis.

There is no reason a priori why β should depend on σ and $\underset{\sim}{\phi}$. Therefore we assume that the prior distribution of $(\sigma,\underset{\sim}{\phi},\beta)$ is given by

$$p(\sigma,\underset{\sim}{\phi},\beta) = p(\sigma,\underset{\sim}{\phi})p(\beta) = \sigma^{-1}p(\beta) \ . \qquad (3.4.9)$$

Another useful concept suggested in the literature (Box and Tiao (1973)) is the one of a reference prior for β. This is usually, but not necessarily, taken to be a uniform prior and is intended, as its name implies, for reference purposes. It has the property that if the data are viewed in the light of some other prior distribution the new posterior distribution could be readily obtained by using the reference prior. We find it convenient to use a uniform reference prior for β. Using the prior distribution in (3.4.9) the posterior distribution is given by

$$p(\sigma,\beta,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p) = p(\beta)[\omega(\beta)]^{n-p}\sigma^{-(n-p+1)}$$

$$\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}\} \qquad (3.4.10)$$

Integrating over σ,

$$p(\beta,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p) = p(\beta)[\omega(\beta)]^{n-p}\frac{1+\beta}{2}$$

$$\Gamma(\tfrac{n-p}{2}(1+\beta))[c(\beta)]^{-\frac{n-p}{2}(1+\beta)}$$

$$\{\sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}\}^{-\frac{n-p}{2}(1+\beta)}$$

$$= p(\beta)\frac{\Gamma(1+\frac{n-p}{2}(1+\beta))}{[\Gamma(1+\frac{1}{2}(1+\beta))]^{n-p}}$$

$$\{\sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}\}^{-\frac{n-p}{2}(1+\beta)} \qquad (3.4.1$$

To get the posterior distribution of β we have to integrate over $\underset{\sim}{\phi}$,

$$p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p) = \int_R p(\beta,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p)d\underset{\sim}{\phi} \qquad (3.4.1$$

where R is the stationarity region for the AR(p) process; $R = \{\underset{\sim}{\phi}$: such that $\phi(B) = 0$ has roots outside the unit circle

It is not possible to obtain a closed form expression fo $p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p)$. However for low order autoregressive processes th integral in equation (3.4.12) can be evaluated numerically.

The posterior distribution $p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p)$ serves as weight function in deriving the posterior distribution of $\underset{\sim}{\phi}$.

$$p(\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p) = \int_{-1}^{+1} p(\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta)\ p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p)d\beta \qquad (3.4.1$$

Again, for low order processes this posterior distribution is readily evaluated using numerical integration techniques.

Conditional posterior distribution $p(\phi|\underset{\sim}{z},z_1,\beta)$ for first order autoregressive processes

The conditional posterior distribution of ϕ given β is

$$p(\phi|\underset{\sim}{z},z_1,\beta) \propto \{S(\phi;\beta)\}^{-\frac{n-1}{2}(1+\beta)} \qquad (3.4.14)$$

where: $S(\phi;\beta) = \sum_{t=2}^{n} |z_t-\phi z_{t-1}|^{2/1+\beta}$

The following properties of $p(\phi|\underset{\sim}{z},z_1,\beta)$ are readily derived:

LEMMA: $S(\phi;\beta)$ is convex, continuous, and for $\beta < 1$ has a continuous first derivative. Furthermore $S(\phi;\beta)$ has a unique minimum which is obtained in the interval $[y_{(1)},y_{(n)}]$ where $y_{(i)}$ are the ordered $\frac{z_t}{z_{t-1}}$

Proof:

$$S(\phi;\beta) = \sum_{t=2}^{n} s_t(\phi;\beta) \text{ where: } s_t(\phi;\beta) = |z_t-\phi z_{t-1}|^{\frac{2}{1+\beta}}$$

$s_t(\phi;\beta)$ is continuous and for $\beta < 1$ ($\Leftrightarrow \frac{2}{1+\beta} > 1$) the derivative of $s_t(\phi;\beta)$ exists and is given by

$$s_t'(\phi;\beta) = \frac{2}{1+\beta} |z_t-\phi z_{t-1}|^{\frac{2}{1+\beta}-1} |z_{t-1}|\delta_t \qquad (3.4.15)$$

where

$$\delta_t = \begin{cases} -1 \text{ if } \frac{z_t}{z_{t-1}} > \phi \\ +1 \text{ if } \frac{z_t}{z_{t-1}} < \phi \end{cases}$$

When ϕ goes to $\frac{z_t}{z_{t-1}}$ the right and the left hand limit of $s_t'(\phi;\beta)$ coincide, thus proving continuity of the first derivative. Furthermore for $-1 < \beta \leq 1$

$$s_t(\frac{\phi_1+\phi_2}{2};\beta) = \left|z_t - \frac{\phi_1+\phi_2}{2} z_{t-1}\right|^{\frac{2}{1+\beta}}$$

$$= \left|\frac{z_t-\phi_1 z_{t-1}}{2} + \frac{z_t-\phi_2 z_{t-1}}{2}\right|^{\frac{2}{1+\beta}}$$

$$\leq \left[\frac{|z_t-\phi_1 z_{t-1}|}{2} + \frac{|z_t-\phi_2 z_{t-1}|}{2}\right]^{\frac{2}{1+\beta}}$$

$$\leq \frac{|z_t-\phi_1 z_{t-1}|^{2/1+\beta}}{2} + \frac{|z_t-\phi_2 z_{t-1}|^{2/1+\beta}}{2}$$

$$= \frac{s_t(\phi_1;\beta)+s_t(\phi_2;\beta)}{2}$$

thus proving convexity. From (3.4.15) we also see that $s_t'(\phi;\beta)$ is a monotonically increasing function of ϕ.

This proves that $S(\phi;\beta) = \sum_{t=2}^{n} s_t(\phi;\beta)$ is convex, continuous and for $\beta < 1$ has a continuous derivative. Furthermore, the fact that $S'(\phi;\beta) = \frac{\partial S(\phi;\beta)}{\partial\phi}$ is monotonically increasing, assures that the derivative can vanish once and only once, and that the extreme value must be a minimum.

Q.E.D.

In the case $\beta = 0$ the posterior distribution of ϕ is given by

$$p(\phi|\underline{z},z_1,\beta=0) = \{\sum_{t=2}^{n}(z_t-\phi z_{t-1})^2\}^{-\frac{n-1}{2}}$$

$$= \{1 + \frac{\sum_{t=2}^{n} z_{t-1}^2(\phi-\hat{\phi})^2}{\sum_{t=2}^{n}(z_t-\hat{\phi}z_{t-1})^2}\}^{-\frac{n-1}{2}} \qquad (3.4.16)$$

This is a $t(\hat{\phi},r_{\hat{\phi}}^2,n-2)$ distribution with mean $\hat{\phi} = \frac{\sum z_t z_{t-1}}{\sum z_{t-1}^2}$ and variance $r_{\hat{\phi}}^2 = \frac{1}{n-2}\frac{\sum z_t^2}{\sum z_{t-1}^2}[1 - \frac{(\sum z_t z_{t-1})^2}{(\sum z_t^2)(\sum z_{t-1}^2)}]$ and $n-2$ degrees of freedom.

The form $t(\hat{\phi},a,v)$ is a logical choice for an approximating distribution for $p(\phi|\underline{z},z_1,\beta)$ in the case $\beta \neq 0$. Lund (1967) suggested this type of approximation for the posterior distribution of the mean of independent observations. His work can readily be applied to the case of a first order autoregressive process. His ideas suggest expanding $S(\phi;\beta)$ around its mode:

$$S(\phi;\beta) = S(\hat{\phi};\beta) + d(\phi-\hat{\phi})^2 \qquad (3.4.17)$$

Using this expansion, the posterior distribution $p(\phi|\underline{z},z_1,\beta)$ in (3.4.14) can be approximated by a $t(\hat{\phi},r_{\hat{\phi}}^2,(n-1)(1+\beta)-1)$ distribution where $\hat{\phi}$ is the mode of (3.4.14) and

$$r_{\hat{\phi}}^2 = \frac{S(\hat{\phi};\beta)}{(n-1)(1+\beta)-1}\frac{1}{d}$$

Various methods for determining the mode $\hat{\phi}$ and the constant d can be employed and they are in a different context discussed in Lund's thesis. Special care is required for the case $\beta > 0$, since in this case the second derivative of $S(\phi;\beta)$ does not exist.

A quick and easy method to derive the mode depends on the fact proved before, that $S(\phi;\beta)$ has a unique minimum and that its derivative $S'(\phi;\beta) = \frac{2}{1+\beta}\sum_{t=2}^{n}|z_t - z_{t-1}|^{\frac{2}{1+\beta}-1}|z_{t-1}|\delta_t$

$$\text{where } \delta_t = \begin{cases} -1 & \text{if } \frac{z_t}{z_{t-1}} > \phi \\ +1 & \text{if } \frac{z_t}{z_{t-1}} < \phi \end{cases}$$

is monotonically increasing.

Calculating $S'(\phi;\beta)$ at some point ϕ, one knows in which direction to proceed towards the mode $\hat{\phi}$. Increasing (decreasing) the value of ϕ until the sign of $S'(\phi;\beta)$ changes and repeating this process with shrinking stepsizes will determine $\hat{\phi}$ to any desired accuracy.

Another method which can be employed to determine the [o]de $\hat{\phi}$ is Newton's iteration procedure. Using this technique [i]t can be shown that $\hat{\phi}$ is a weighted average of cross [p]roducts,

$$\hat{\phi} = \sum_{t=2}^{n} v_t (z_t z_{t-1})$$

[wh]ere $v_t = \dfrac{|z_t - \hat{\phi} z_{t-1}|^{-\frac{2\beta}{1+\beta}}}{\sum_{k=2}^{n} |z_k - \hat{\phi} z_{k-1}|^{-\frac{2\beta}{1+\beta}} z_{k-1}^2}$

[H]owever, when $\beta > 0$ this procedure has to be used with great [c]aution since several of the terms $|z_t - \phi z_{t-1}|^{-\frac{2\beta}{1+\beta}}$ could [b]ecome infinite.

For $\beta < 0$, the second derivative of $S(\phi;\beta)$ exists and [d] could be derived from the Taylor series expansion.

$$d = \frac{1}{2} \frac{\partial^2 S(\phi;\beta)}{\partial \phi^2}\Big|_{\hat{\phi}} = \frac{1}{2} \frac{1-\beta}{(1+\beta)^2} \sum_{t=2}^{n} |z_t - \hat{\phi} z_{t-1}|^{\frac{-2\beta}{1+\beta}} z_{t-1}^2$$

[F]or $\beta > 0$ it can be chosen such that $[S(\phi_j;\beta) - (S(\hat{\phi},\beta) + d(\phi_j - \hat{\phi})^2)]^2$ is minimized.

[3].5 Parameter estimation for moving average models of order q with shocks from the family of symmetric exponential power distributions.

We consider the invertible model

$$z_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \qquad (3.5.1)$$

where z_t is a stationary difference of original observations with $Ez_t = 0$ and where the a_t are independent drawings from the distribution given in (3.3.2).

To assure invertibility we assume that the roots of $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = 0$ lie outside the unit circle.

$$p(a_1, a_2, \dots, a_n | \sigma, \beta) = [\omega(\beta)]^n \sigma^{-n} \exp\left(-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=1}^{n} |a_t|^{\frac{2}{1+\beta}}\right) \qquad (3.5.2)$$

Making the transformation:

$$z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \qquad 1 \le t \le n$$

and treating the starting values $\underline{a}_*' = (a_0, a_{-1}, \dots, a_{-(q-1)})$ as additional nuisance parameters, we get

$$p(\underline{z} | \sigma, \beta, \underline{\theta}, \underline{a}_*) = [\omega(\beta)]^n \sigma^{-n} \exp\left(-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=1}^{n} \left| \sum_{j=0}^{t-1} \pi_j z_{t-j} + \pi_t a_0 + (\theta_2 \pi_{t-1} + \dots + \theta_q \pi_{t-(q-1)}) a + \dots + \theta_q \pi_{t-1} a_{-(q-1)} \right|^{\frac{2}{1+\beta}}\right) \qquad (3.5.3)$$

where: $\underline{z}' = (z_1, z_2, \dots, z_n)$

$\underline{\theta}' = (\theta_1, \dots, \theta_q)$

$\pi_0 = -1$ and the π_j-weights $(j \ge 1)$ are the coefficients in the expansion

$$\pi(B) = 1 - \sum_{j=0}^{\infty} \pi_j B^j = (1-\theta_1 B - \dots - \theta_q B^q)^{-1}$$

and are given by:

$$\begin{cases} \pi_j = \theta_1 \pi_{j-1} + \dots + \theta_{j-1}\pi_1 - \theta_j & \text{for } 1 \le j \le q \\ \pi_j = \theta_1 \pi_{j-1} + \dots + \theta_q \pi_{j-q} & \text{for } j \ge q+1 \end{cases}$$

The likelihood function for the model in (3.5.1) is given by

$$\ell(\sigma,\beta,\underset{\sim}{\theta},\underset{\sim}{a}_*|\underset{\sim}{z}) = [\omega(\beta)]^n \sigma^{-n}$$

$$\exp\Big(-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=1}^{n} \Big|\sum_{j=0}^{t-1} \pi_j z_{t-j} + \pi_t a_0 + (\theta_2 \pi_{t-1} + \dots + \theta_q \pi_{t-(q-1)}) a_{-1}$$

$$+ \dots + \theta_q \pi_{t-1} a_{-(q-1)}\Big|^{\frac{2}{1+\beta}} \qquad (3.5.4)$$

First we consider the analysis of the MA(q) model for a fixed symmetric parent, thus treating β as given.

Combining the prior distribution

$$p(\sigma,\underset{\sim}{\theta},\underset{\sim}{a}_*) \propto \sigma^{-1} \qquad (3.5.5)$$

with the likelihood in (3.5.4), we derive the posterior distribution of $(\sigma,\underset{\sim}{\theta},\underset{\sim}{a}_*)$ for given β

$$p(\sigma,\underset{\sim}{\theta},\underset{\sim}{a}_*|\underset{\sim}{z},\beta) \propto \sigma^{-(n+1)}$$

$$\exp\Big(-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=1}^{n} \Big|\sum_{j=0}^{t-1} \pi_j z_{t-j} + \pi_t a_0 + (\theta_2 \pi_{t-1} + \dots + \theta_q \pi_{t-(q-1)}) a_{-1}$$

$$+ \dots + \theta_q \pi_{t-1} a_{-(q-1)}\Big|^{\frac{2}{1+\beta}}\Big) \qquad (3.5.6)$$

Integrating over the parameter σ yields

$$p(\underset{\sim}{\theta},\underset{\sim}{a}_*|\underset{\sim}{z},\beta) \propto$$

$$\Big\{\sum_{t=1}^{n} \Big|\sum_{j=0}^{t-1} \pi_j z_{t-j} + \pi_t a_0 + (\theta_2 \pi_{t-1} + \dots + \theta_q \pi_{t-(q-1)}) a_{-1}$$

$$+ \dots + \theta_q \pi_{t-1} a_{-(q-1)}\Big|^{\frac{2}{1+\beta}}\Big\}^{-\frac{n}{2}(1+\beta)} \qquad (3.5.7)$$

In order to derive the conditional posterior distribution $p(\underset{\sim}{\theta}|\underset{\sim}{z},\beta)$ we have to integrate over the starting values $\underset{\sim}{a}_*$.

$$p(\underset{\sim}{\theta}|\underset{\sim}{z},\beta) = \int p(\underset{\sim}{\theta},\underset{\sim}{a}_*|\underset{\sim}{z},\beta)\, d\underset{\sim}{a}_* \qquad (3.5.8)$$

In practice this integration is not easy to do. However we know that for any invertible process $\pi(B)$ converges for $|B| \le 1$. Thus, if we work with a moderate number of observations and if the moving average parameters $\underset{\sim}{\theta}$ satisfy the invertibility condition, the π-weights in (3.5.7) will die out rather rapidly and the choice of the starting values will not be critical.

An approximation to the integral in (3.5.8) is given by substituting the mode of $\underset{\sim}{a}_*$ into equation (3.5.7). One then has to find the choice of $\underset{\sim}{a}_*$ which minimizes

$$\sum_{t=1}^{n} \Big|\sum_{j=0}^{t-1} \pi_j z_{t-j} + \pi_t a_0 + (\theta_2 \pi_{t-1} + \dots + \theta_q \pi_{t-(q-1)}) a_{-1}$$

$$+ \dots + \theta_q \pi_{t-1} a_{-(q-1)}\Big|^{\frac{2}{1+\beta}}$$

For the case $\beta = 0$, this will be the usual least squares estimate, and is discussed in Box-Jenkins (1970). For general β this minimization problem can be solved by the use of nonlinear minimization routines.

Another sensible approximation to the integral in (3.5.8), and one which is more convenient in practice, is to approximate the posterior distribution $p(\underset{\sim}{\theta}|\underset{\sim}{z},\beta)$ by setting the starting values in (3.5.7) equal to their expected values.

$$p(\underset{\sim}{\theta}|\underset{\sim}{z},\beta) = \int p(\underset{\sim}{\theta},\underset{\sim}{a}_*|\underset{\sim}{z},\beta)\,d\underset{\sim}{a}_* = p(\underset{\sim}{\theta},\underset{\sim}{a}_*=\underset{\sim}{0}|\underset{\sim}{z},\beta)$$
$$= \left\{\sum_{t=1}^{n}\left|\sum_{j=0}^{t-1}\pi_j z_{t-j}\right|^{\frac{2}{1+\beta}}\right\}^{-\frac{n}{2}(1+\beta)} \qquad (3.5.9)$$

When $\underset{\sim}{\theta}$ approaches the non invertibility region more refined methods in determining the starting values will have to be used. For the examples of section 3.7 this will not be necessary since the parameter θ in the considered MA(1) model is well within the invertibility region, and the π-weights die out rapidly.

The conditional posterior distribution of $\underset{\sim}{\theta}$ given in equation (3.5.9) shows how the inference about $\underset{\sim}{\theta}$ changes with different assumptions about the parent distribution in the probability model.

If we treat β as a random variable with prior distribution $p(\beta)$, we obtain the posterior distribution

$$p(\sigma,\beta,\underset{\sim}{\theta}|\underset{\sim}{z}) = p(\sigma,\beta,\underset{\sim}{\theta},\underset{\sim}{a}_*=\underset{\sim}{0}|\underset{\sim}{z}) = p(\beta)[\omega(\beta)]^n\sigma^{-(n+1)}$$
$$\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\sum_{t=1}^{n}\left|\sum_{j=0}^{t-1}\pi_j z_{t-j}\right|^{\frac{2}{1+\beta}}\right\} \qquad (3.5.10)$$

Integrating over the nuisance parameter σ, we get

$$p(\beta,\underset{\sim}{\theta}|\underset{\sim}{z}) = p(\beta)\,\frac{\Gamma(1+\frac{n}{2}(1+\beta))}{[\Gamma(1+\frac{1}{2}(1+\beta))]^n}$$
$$\left\{\sum_{t=1}^{n}\left|\sum_{j=0}^{t-1}\pi_j z_{t-j}\right|^{\frac{2}{1+\beta}}\right\}^{-\frac{n}{2}(1+\beta)} \qquad (3.5.11)$$

We further integrate over the moving average parameter $\underset{\sim}{\theta}$ to derive the posterior distribution of β.

$$p(\beta|\underset{\sim}{z}) = \int_R p(\beta,\underset{\sim}{\theta}|\underset{\sim}{z})\,d\underset{\sim}{\theta} \qquad (3.5.12)$$

where R is the invertibility region of the MA(q) process.

As in the case of autoregressive processes it is not possible to derive a closed form for the posterior distribution in (3.5.12). However, if the order of the MA process is low, integration can be carried out using numerical integration methods.

The posterior distribution of β acts like a weight function in deriving the posterior distribution of $\underset{\sim}{\theta}$,

$$p(\underset{\sim}{\theta}|\underset{\sim}{z}) = \int_{-1}^{1} p(\underset{\sim}{\theta}|\underset{\sim}{z},\beta)\, p(\beta|\underset{\sim}{z})\,d\beta \qquad (3.5.13)$$

3.6. Forecasting time series models when the shocks are from the family of symmetric exponential power distributions.

The first order integrated moving average model and its one step ahead predictive distribution.

In the examples of section 3.7 we will discuss integrated moving average models of the form $z_t - z_{t-1} = a_t - \theta a_{t-1}$. This type of model is particularly important in practice, since many economic, business and engineering data behave approximately according to this model. Furthermore, as pointed out in section 3.2, stock price data follow a model of this kind in which the moving average parameters is close to zero.

Sample theory approach to forecasting:

We distinguish two approaches to forecasting. The first one is a sampling theory approach. As discussed in section 1.4, the minimum mean squared error forecast of the future observation $z_{n+\ell}$, given all the observations up to time n, is the conditional expectation of $z_{n+\ell}$ at time n. For any class of distributions for which the variance exists, the distributional assumptions about the shocks a_t are irrelevant for the derivation of the minimum mean squared error forecast. Forecasts are of little value if they

are not accompanied by some measure of their variability. The variance of the forecast error

$$e_n(\ell) = z_{n+\ell} - \hat{z}_n(\ell) = a_{n+\ell} + \sum_{j=1}^{\ell-1} \psi_j a_{n+\ell-j} \qquad (3.6.1)$$

provides such a measure, and it is given by $V(e_n(\ell)) = \sigma_a^2(1+\sum_{j=1}^{\ell-1}\psi_j^2)$. The distributional assumptions about the shocks a_t change the probabilistic interpretation of the probability interval

$$\{\hat{z}_n(\ell) \pm \lambda[V\,(e_n(\ell))]^{\frac{1}{2}}\} \qquad (3.6.2)$$

If one is interested in one step ahead forecasts, the forecast error is given by $e_n(1) = a_{n+1}$. Thus, α probability limits for the future observation z_{n+1} are given by $\{\hat{z}_n(1) \pm \lambda\sigma_a\}$ where λ is chosen such that $\text{Prob}\{|a|>\lambda\sigma_a\} = \alpha$. For the case $\beta = 0$, the Normal table provides the corresponding α and λ. If $\beta > 0$ the distribution of the shocks a_t is leptokurtic and the Normal probability limits will underestimate the risk of a realization in the extreme tails. For the case $\beta < 0$ the Normal theory probability limits will overestimate this risk, since the distribution is platykurtic. Box-Tiao (1973) show that the probability limits can be quite different for very small $\alpha(\alpha<<.05)$; however for $\alpha = .05$ the probability limits are not very sensitive to the choice of β.

For lead times $\ell \geq 2$ the distribution of the forecast errors given in (3.6.1) can be readily derived in the case of a Normal parent. For general β, however, the distribution of $e_n(\ell)$ is very complicated. Nevertheless, some idea of the approach to Normality of $e_n(\ell)$ can be obtained by considering the measure of kurtosis of the forecast errors.

It is readily shown that the kurtosis of the ℓ-step ahead forecast error $e_n(\ell) = a_{n+\ell}+\psi_1 a_{n+\ell-1}+\cdots+\psi_{\ell-1}a_{n+1}$ is given by

$$\gamma_2(e_n(\ell)) = \frac{1+\sum_{j=1}^{\ell-1}\psi_j^4}{[1+\sum_{j=1}^{\ell-1}\psi_j^2]^2}\,\gamma_2(a) \qquad (3.6.3)$$

The kurtosis of the distribution of the ℓ-step ahead forecast error depends on

i) the non Normality parameter in the distribution of the shocks

ii) the ψ-weights of the ARIMA model

For the case of an ARIMA (0,1,1) model, the ψ-weights are given by

$$\psi_j = (1-\theta) \quad \text{for all } j \geq 1$$

and

$$\gamma_2(e_n(\ell)) = \frac{1+(\ell-1)(1-\theta)^4}{[1+(\ell-1)(1-\theta)^2]^2}\,\gamma_2(a) \qquad (3.6.4)$$

In Figure 3.2 we plot $\frac{\gamma_2(e_n(\ell))}{\gamma_2(a)}$ for various lead times ℓ to illustrate the amount of non Normality of the distribution of $e_n(\ell)$.

Since the ARIMA (0,1,1) process is non stationary, the ψ-weights will not die out, and for moderate lead times the distribution of the forecast errors can be approximated by a Normal distribution.

The above sample theory interpretation of forecasting has the drawback that it assumes that the values of the parameters are known. Parameters, however, are estimated and in practice there are parameter estimation errors present. A sample theory development which allowed for errors in the parameters would be extremely difficult. However some progress has been made by investigating how much the variance of the forecast errors increases if the parameters are estimated from data. Some work of this kind has been done by Box and Jenkins (1970) and by Bloomfield (1972). We shall here consider a different approach.

Bayesian approach to forecasting:

Another approach to forecasting is a Bayesian one. This approach does provide a managable way of incorporating errors in the parameters. Treating the parameters in the ARIMA

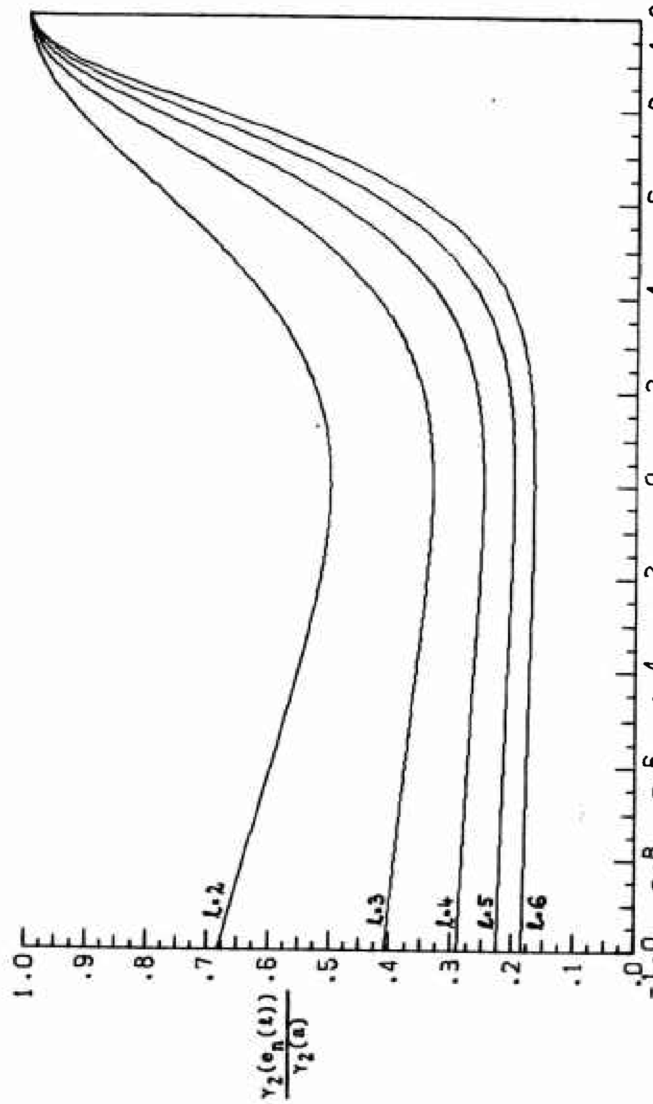


Figure 3.2: Plot of $\frac{\gamma_2(e_n(\ell))}{\gamma_2(a)}$ for several lead times ℓ: ARIMA (0,1,1) model.

model as random variables the predictive distribution of the future observation $z_{n+\ell}$ given the data up to time n is derived. Wichern (1969) pursued this approach for the case of Normal shocks.

Even though the following results could be extended to the general ARIMA (p,d,q) model, we will only consider the case of the integrated first order moving average process. Furthermore we will only discuss the one step ahead predictive distribution.

<u>The predictive distribution $p(z_{n+1}|\underset{\sim}{z})$ for integrated first order moving average models with errors from the class of symmetric exponential power distributions</u>

$$p(z_{n+1}|\underset{\sim}{z}) = \int_{-1}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty} p(z_{n+1},\sigma,\beta,\theta|\underset{\sim}{z})d\sigma\, d\beta\, d\theta \quad (3.6.5)$$

where $\underset{\sim}{z}' = (z_1,z_2,\ldots,z_n)$.

$$p(z_{n+1},\sigma,\beta,\theta|\underset{\sim}{z}) = p(z_{n+1}|\sigma,\beta,\theta,\underset{\sim}{z})\; p(\sigma,\beta,\theta|\underset{\sim}{z}) \quad (3.6.6)$$

and

$$p(z_{n+1}|\sigma,\beta,\theta,\underset{\sim}{z}) = \omega(\beta)\sigma^{-1}$$

$$\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\,|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}\} \quad (3.6.7)$$

where $\hat{z}_n(1) = (1-\theta)\sum_{j=0}^{n-1}\theta^j z_{n-j}$

In equation (3.5.10) we derived the posterior distribution

$$p(\sigma,\beta,\theta|\underset{\sim}{z}) \propto p(\beta)\ [\omega(\beta)]^{n}\sigma^{-(n+1)}$$

$$\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\, S(\theta;\beta)\} \qquad (3.6.8)$$

where

$$S(\theta;\beta) = \sum_{t=1}^{n}\left|\sum_{j=0}^{t-1}\pi_j w_{t-j}\right|^{\frac{2}{1+\beta}}$$

The w_t are the first difference of the original observations and $\pi_0 = -1$ and $\pi_j = -\theta^j \quad j \geq 1$

Substituting (3.6.7) and (3.6.8) into (3.6.6) yields

$$p(z_{n+1},\sigma,\beta,\theta|\underset{\sim}{z}) \propto p(\beta)\ [\omega(\beta)]^{n+1}\sigma^{-(n+2)}$$

$$\exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\, [|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}+S(\theta;\beta)]\} \qquad (3.6.9)$$

Integrating over σ we get the posterior distribution

$$p(z_{n+1},\beta,\theta|\underset{\sim}{z}) \propto p(\beta)\ \frac{\Gamma(1+\frac{n+1}{2}(1+\beta))}{[\Gamma(1+\frac{1}{2}(1+\beta))]^{n+1}}$$

$$\{|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}+S(\theta;\beta)\}^{-\frac{n+1}{2}(1+\beta)} \qquad (3.6.10)$$

It is of interest to consider the conditional predictive distribution for various choices of the parent distribution;

$$p(z_{n+1},\theta|\underset{\sim}{z},\beta) \propto \{|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}+S(\theta;\beta)\}^{-\frac{n+1}{2}(1+\beta)}$$

$$\propto [S(\theta;\beta)]^{-\frac{1}{2}(1+\beta)}$$

$$\{1+\frac{|z_{n+1}-\hat{z}_n(1)|^{2/1+\beta}}{S(\theta;\beta)}\}^{-\frac{n+1}{2}(1+\beta)}\quad p(\theta|\underset{\sim}{z},\beta) \qquad (3.6.11)$$

where $p(\theta|\underset{\sim}{z},\beta)$ is the posterior distribution of θ for a fixed parent distribution, given in (3.5.9). Therefore, the posterior predictive distribution for given β and θ is given by

$$p(z_{n+1}|\underset{\sim}{z},\beta,\theta) \propto \{1+\frac{|z_{n+1}-\hat{z}_n(1)|^{2/1+\beta}}{S(\theta;\beta)}\}^{-\frac{n+1}{2}(1+\beta)} \qquad (3.6.12)$$

and the posterior predictive distribution of z_{n+1} for a specific parent by

$$p(z_{n+1}|\underset{\sim}{z},\beta) \propto \int_{-1}^{1} p(z_{n+1}|\underset{\sim}{z},\beta,\theta)\ p(\theta|\underset{\sim}{z},\beta)d\theta \qquad (3.6.13)$$

In the case $\beta = 0$, $\frac{z_{n+1}-\hat{z}_n(1)}{S(\theta;\beta)}$ follows a t-distribution with n degrees of freedom and $p(z_{n+1}|\underset{\sim}{z},\beta=0)$ is an average of t-distributions, weighted by $p(\theta|\underset{\sim}{z},\beta=0)$. Equation (3.6.13) allows one to determine the sensitivity (or conversely the robustness) of the predictive distribution to changes in the assumptions about the distribution of the shocks a_t.

$p(z_{n+1}|\underset{\sim}{z})$ is derived by averaging the conditional predictive distributions in (3.6.13). The posterior distribution of β given in (3.5.12) acts as the weighting function.

$$p(z_{n+1}|\underset{\sim}{z}) = \int_{-1}^{1} p(z_{n+1}|\underset{\sim}{z},\beta)\ p(\beta|\underset{\sim}{z})d\beta \qquad (3.6.14)$$

3.7 Examples

Rather than generating artificial data to illustrate the above results we shall consider actually observed data. Two series are considered. One was collected at a chemical plant (bihourly chemical process concentration readings given on page 525 of Box and Jenkins (1970)). The other is an economic series (daily IBM common stock closing prices, given on page 526 of Box and Jenkins (1970) and already discussed in a different context in chapter 2 of this thesis).

Example 3.1: Series A: bihourly chemical process concentration readings

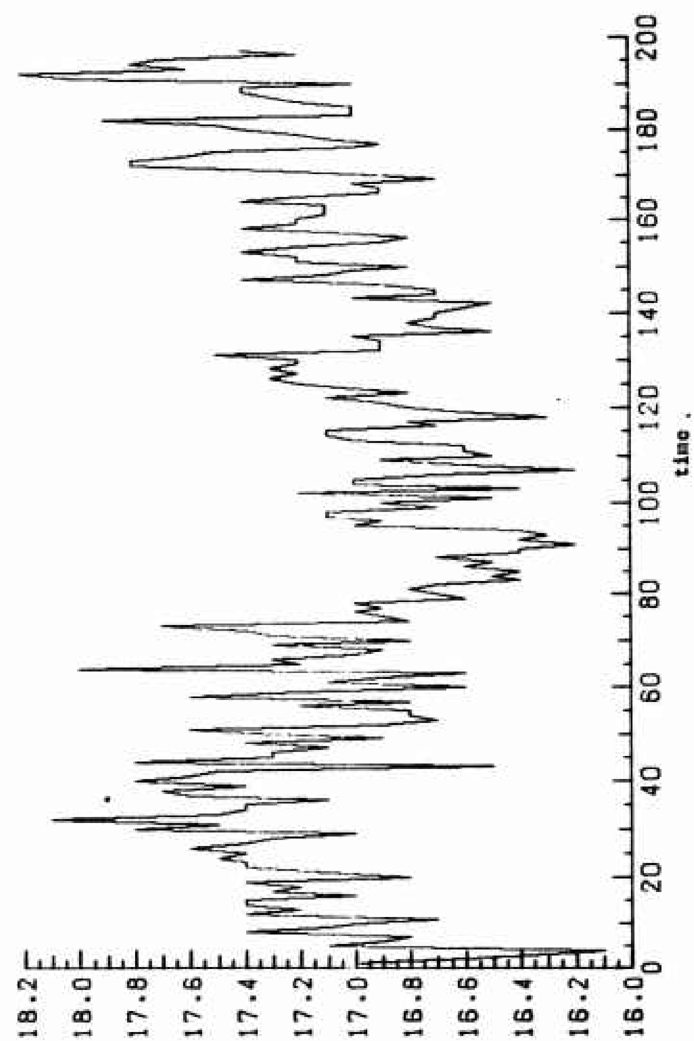
Series A is plotted in Figure 3.3. Identification, estimation under the assumption of Normal distributed errors and diagnostic checking for series A are carried out in Box and Jenkins. The entertained model is given by an integrated first order moving average model

$$z_t - z_{t-1} = a_t - \theta a_{t-1} \qquad (3.7.1)$$

Assuming Normal shocks a_t, the point estimates of the parameters are

$$\hat{\theta} = .7 \quad \text{and} \quad \hat{\sigma}_a^2 = .101$$

We broaden the model (3.7.1) by introducing the parameter β of the family of exponential power distributions, thus



allowing for greater flexibility in the distribution of the shocks a_t. In the following we assume a uniform reference prior for $\beta(-1<\beta\leq 1)$.

The posterior probability distribution $p(\beta,\theta|\underset{\sim}{z})$, which is given in (3.5.11), is derived and its contours are given in Figure 3.4. The density levels of the two outer contours are

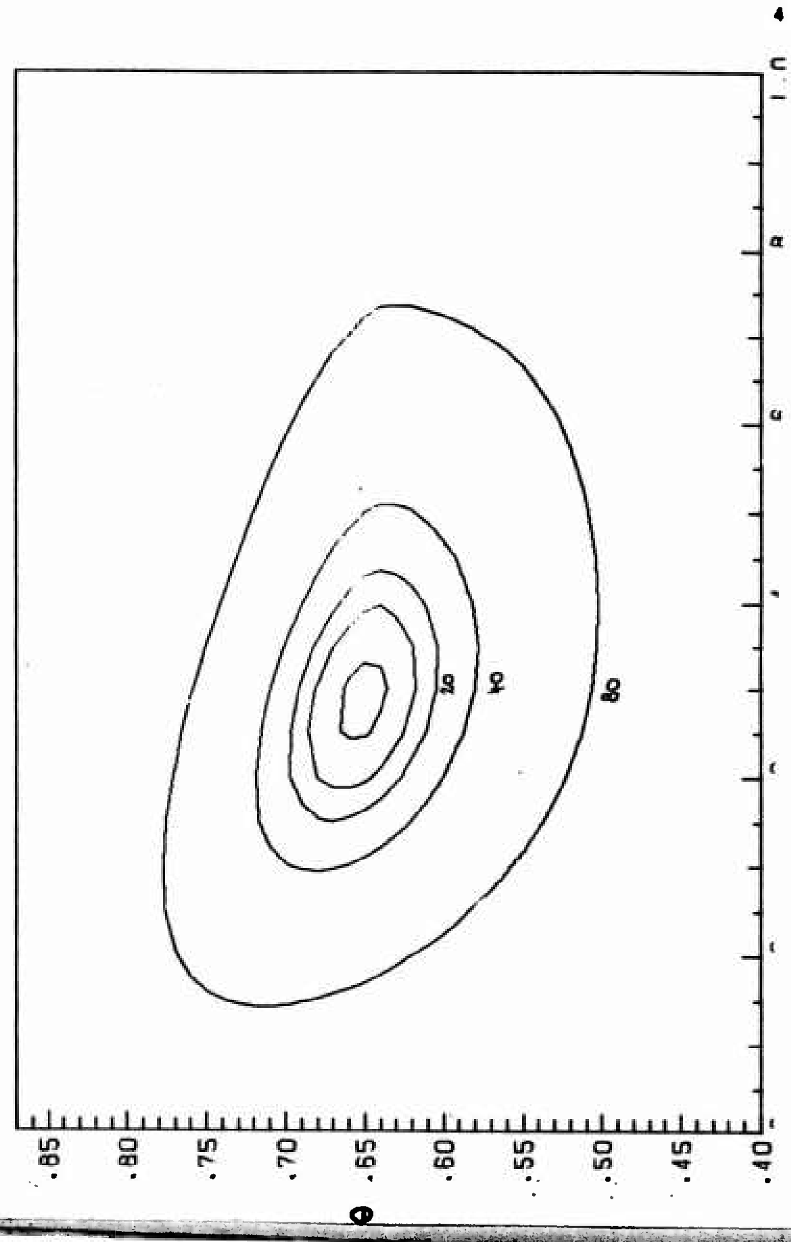
$$p(\beta,\theta|\underset{\sim}{z}) = .2\ p(\hat{\beta},\hat{\theta}|\underset{\sim}{z})$$
$$p(\beta,\theta|\underset{\sim}{z}) = .6\ p(\hat{\beta},\hat{\theta}|\underset{\sim}{z})$$

where $(\hat{\beta},\hat{\theta})$ is the mode of the posterior distribution $p(\beta,\theta)$. Using the bivariate Normal approximation these contours would very roughly be the boundaries of the 80 and 40 percent highest posterior density regions.

The conditional posterior distributions of θ, $p(\theta|\underset{\sim}{z},\beta)$ of equation (3.5.9), are plotted in Figure 3.5. Two things are worth noting:

i) the conditional inference about θ is somewhat sensitive to changes in the parameter β of the parent distribution of the shocks. The mode of $p(\theta|\underset{\sim}{z},\beta)$ changes over a range from .63 (for $\beta=.9$) to .85 (for $\beta=-.5$),

ii) the conditional posterior distributions $p(\theta|\underset{\sim}{z},\beta)$ are skewed. For the case $\beta = 0$ (Normal shocks)



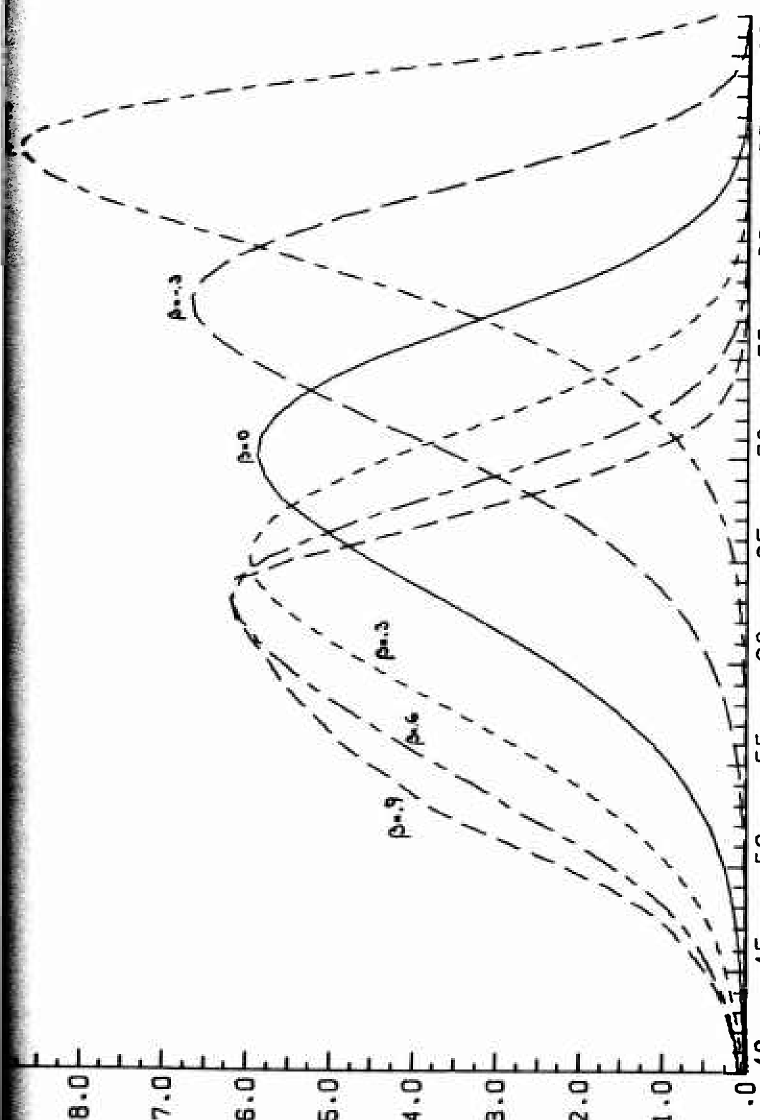


Figure 3.5: Posterior distribution of θ for various fixed β: Series A.

this fact was pointed out by Wichern (1969), who considered several normalizing transformations.

To derive the posterior distribution of the moving average parameter θ the conditional distributions $p(\theta|\underset{\sim}{z},\beta)$ have to be weighted by the posterior distribution $p(\beta|\underset{\sim}{z})$. Assuming a uniform reference prior for β the posterior distribution $p(\beta|\underset{\sim}{z})$ is given in Figure 3.6. This indicates the possibility of a slight leptokurtic distribution for the shocks although the deviation from Normality is probably not large.

The posterior distribution $p(\theta|\underset{\sim}{z})$, given in (3.5.13), and $p(\theta|\underset{\sim}{z},\beta=0)$, the posterior distribution of θ under the usual assumption of Normal distributed shocks, are plotted in Figure 3.7. The two distributions are somewhat different. While the shape of both distributions is similar, the mode occurs of different values (mode of $p(\theta|\underset{\sim}{z})$ occurs at .64, while for $\beta=0$ the mode is given by $\hat{\theta}=.7$).

Two closely linked objectives of time series modelling are forecasting and control. It is of considerable interest to investigate whether the predictive distribution of future observations is affected by the introduction of the parameter β which allows for greater flexibility in the underlying parent distribution of the shocks.

For the ARIMA(0,1,1) model the predicted value of all future times is a constant independent of ℓ. Table 3.1

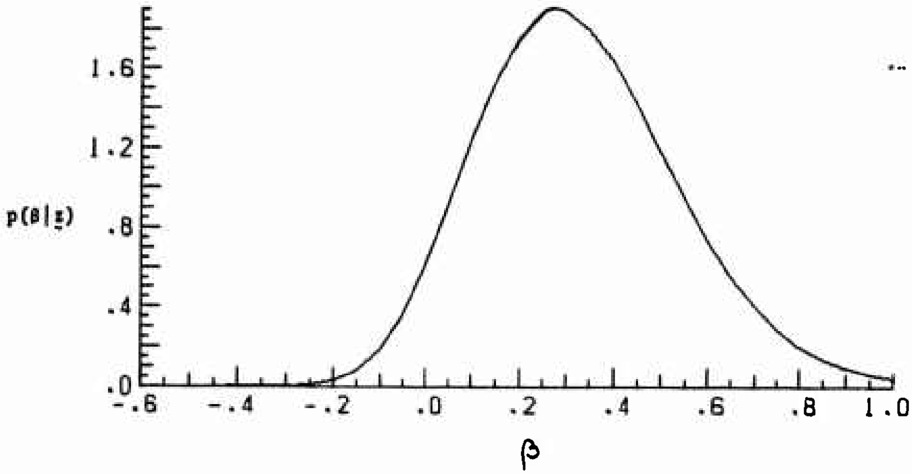


Figure 3.6: Posterior distribution of β assuming a uniform reference prior: Series A.

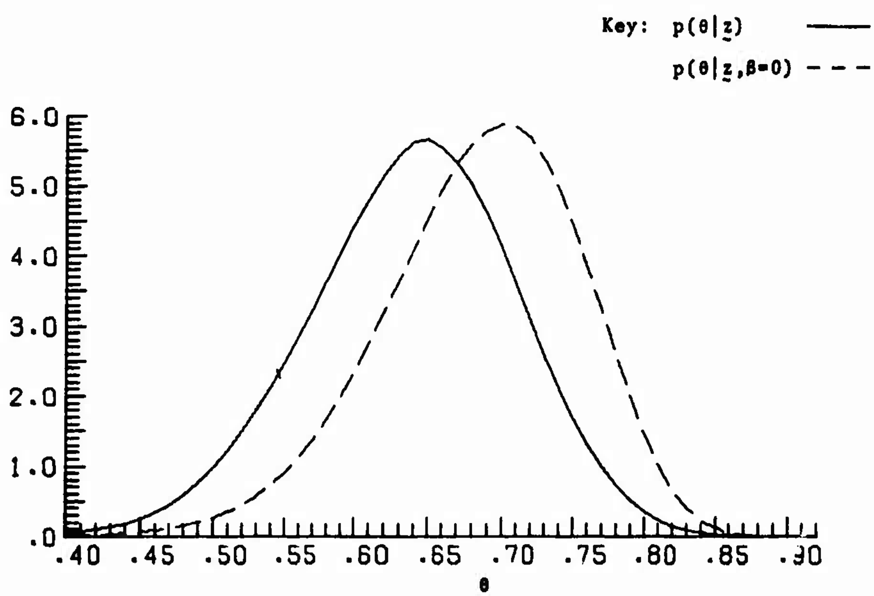


Figure 3.7: Posterior distribution of θ, p(θ|z), and posterior distribution of θ assuming Normal distributed shocks, p(θ|z,β=0): Series A.

gives the predicted values $\hat{z}_n(\ell) = (1-\theta) \sum_{j=0}^{n-1} \theta^j z_{n-j}$ for various values of θ covering the range from $.46 \leq \theta \leq .88$. The forecasts vary from 17.42 to 17.51 which is, compared to the variation of series A, plotted in Figure 3.3, only a very small change. It therefore appears that the forecasts are very insensitive to changes in θ over the range of appreciable posterior probability density $p(\theta|\underset{\sim}{z},\beta)$.

Due to this insensitivity of $z_n(\ell)$ towards changes in θ, the modes of the conditional one step ahead predictive distributions $p(z_{n+1}|\underset{\sim}{z},\beta)$ are similar. The shape of the distribution however will reflect the different assumptions about the parameter β. The conditional one step ahead predictive distributions are plotted in Figure 3.8 for several values of β.

Weighting the different conditional predictive distributions $p(z_{n+1}|\underset{\sim}{z},\beta)$ with the posterior distribution $p(\beta|\underset{\sim}{z})$ leads to the predictive distribution $p(z_{n+1}|\underset{\sim}{z})$. This distribution is plotted in Figure 3.9, together with $p(z_{n+1}|\underset{\sim}{z},\beta=0)$, the one step ahead predictive distribution assuming Normal shocks.

Inspection of Figure 3.9 shows that both distributions are centered around a similar mode. Due to the fact that the posterior distribution of β indicates a slight leptokurtic distribution for the shocks, $p(z_{n+1}|\underset{\sim}{z})$ has more probability in its tails than $p(z_{n+1}|\underset{\sim}{z},\beta=0)$.

θ	$z_t(\ell)$	θ	$z_t(\ell)$
.46	17.42	.68	17.50
.48	17.43	.70	17.50
.50	17.43	.72	17.51
.52	17.44	.74	17.51
.54	17.45	.76	17.51
.56	17.46	.78	17.51
.58	17.46	.80	17.51
.60	17.47	.82	17.51
.62	17.48	.84	17.50
.64	17.49	.86	17.49
.66	17.49	.88	17.47

able 3.1: $\hat{z}_n(\ell) = (1-\theta) \sum_{j=0}^{n-1} \theta^j z_{n+1-j}$ for various θ: Series A (n=197).

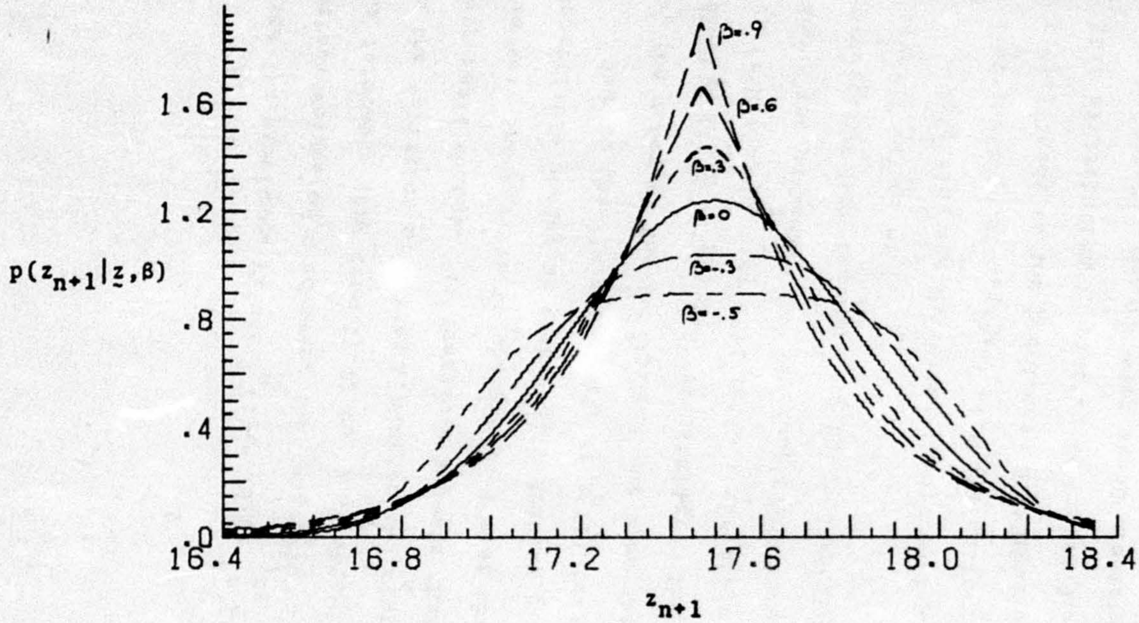


Figure 3.8: Posterior one-step ahead predictive distribution for various fixed β: Series A; $n=197$.

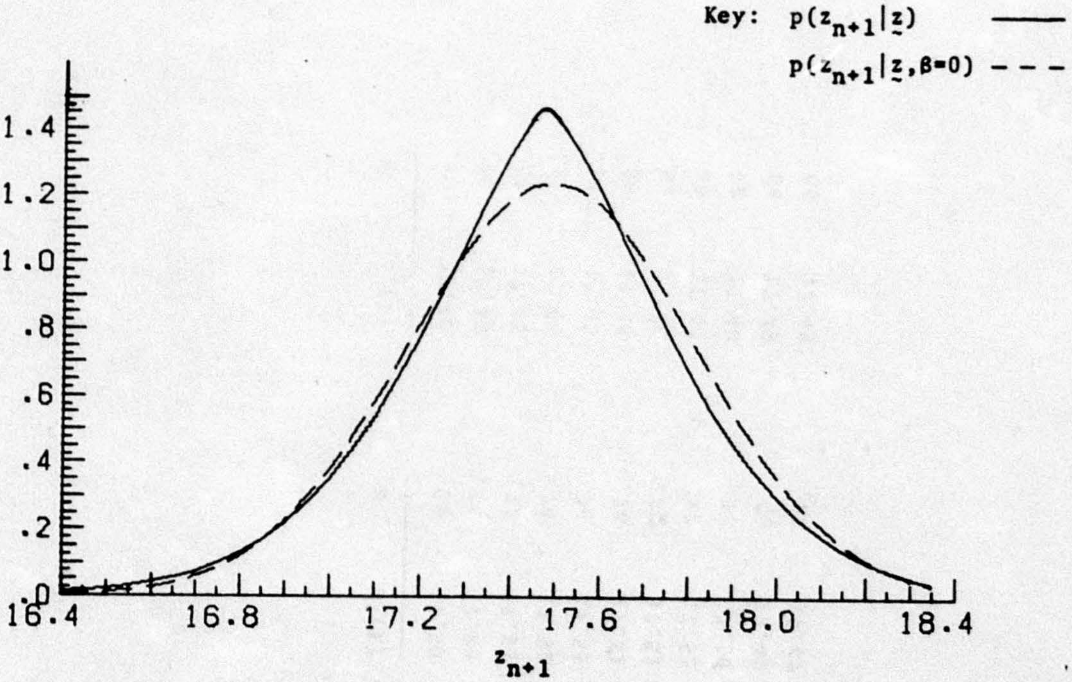


Figure 3.9: Posterior one step ahead predictive distribution $p(z_{n+1}|\underset{\sim}{z})$ compared with $p(z_{n+1}|\underset{\sim}{z},\beta=0)$, the posterior one step ahead predictive distribution assuming Normal distributed shocks: Series A; $n=197$.

Information about the non Normality of the distribution of the forecasts for higher lead times can be derived by considering the kurtosis of the ℓ-step ahead forecast error $e_n(\ell)$ which was given in (3.6.4). In Figure 3.10 we plot the kurtosis

$$\gamma_2(e_n(\ell)) = \frac{1+(\ell-1)(1-\theta)^4}{[1+(\ell-1)(1-\theta)^2]^2} \frac{\Gamma(\frac{5}{2}(1+\beta))\Gamma(\frac{1}{2}(1+\beta))}{[\Gamma(\frac{3}{2}(1+\beta))]^2}$$

for several ℓ, using the mode $(\hat{\beta},\hat{\theta}) = (.3,.65)$ as estimates for β and θ.

Example 3.2: Series B: daily IBM common stock closing prices.

This series is plotted in Figure 2.1. It was shown in Chapter 2 that the model for series B is given by an integrated first order moving average process. Under the assumption of Normal shocks the estimated parameters are

$$\hat{\theta} = -.087 \qquad \hat{\sigma}_a^2 = 52.2$$

Before we derive the posterior distributions for the model with shocks from the symmetric exponential power distribution, we develop a rough preliminary identification tool which indicates whether such an extension is necessary. This preliminary identification tool is similar to the stability test proposed by Fama and Roll (1971) which is critically reviewed in the thesis by Hsu (1973). Furthermore we comment on the question whether the stable Paretian

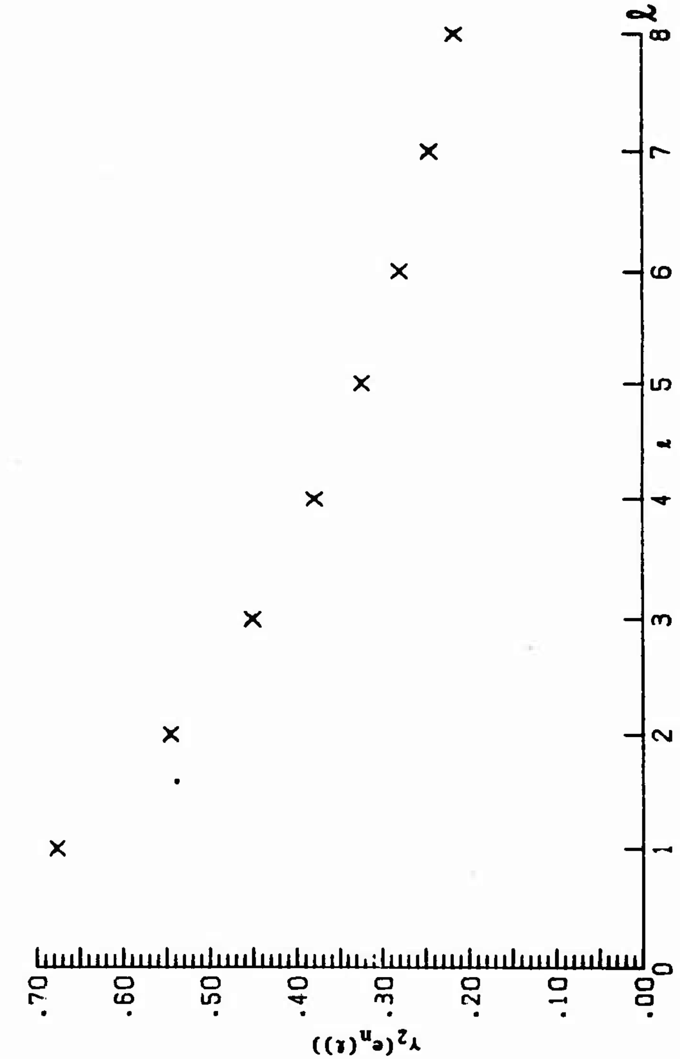


Figure 3.10: Plot of the kurtosis of the ℓ-step ahead forecast error: Series A.

hypothesis, as discussed in section 3.2, provides a sensible alternative hypothesis for the leptokurtic error distribution of series B.

In the following we assume the random walk model, in which $w_t = z_t - z_{t-1}$ is white noise. It follows from the discussion of stable distributions in section 3.2 that if the independent shocks come from a stable distribution with parameters α and γ ($\delta=0;\xi=0$), then

$$w_t^{(k)} = \frac{1}{k^{1/\alpha}}(w_{t+1}+\dots+w_{t+k}) \tag{3.7.3}$$

will have the same distribution as each individual w_t. If, however, the distribution of the shocks has finite moments (for example the family of exponential power distribution) it can be shown that the kurtosis

$$\gamma_2(w_t^{(k)}) = \frac{1}{k}\gamma_2(w_t) \tag{3.7.4}$$

We partition the first differences w_t of series B into m_k non overlapping segments $\{w_1^{(k)}, w_{k+1}^{(k)}, \dots, w_{jk+1}^{(k)}, \dots\}$ and calculate the sample kurtosis defined by

$$r_2(w_t^{(k)}) = \frac{m_k \sum_{j=0}^{m_k-1} (w_{jk+1}^{(k)})^4}{[\sum_{j=0}^{m_k-1} (w_{jk+1}^{(k)})^2]^2} - 3 \tag{3.7.5}$$

If the stable hypothesis were true, $w_t^{(k)}$ would follow the same stable distribution as w_t (for all $k \geq 1$), and the kurtosis of $w_t^{(k)}$ would not exist. If the assumption of Normal shocks were true, the kurtosis of $w_t^{(k)}$ would be zero. If the shocks came from the family of symmetric exponential power distributions the kurtosis would decay according to

$$\gamma_2(w_t^{(k)}) = \frac{1}{k}\left\{\frac{\Gamma(\frac{5}{2}(1+\beta))\Gamma(\frac{1}{2}(1+\beta))}{[\Gamma(\frac{3}{2}(1+\beta))]^2} - 3\right\}$$

Figure 3.11 should be treated as a rough preliminary identification guide only. Visual inspection of the graph of the sample kurtosis of $w_t^{(k)}$ versus k can give some indication whether such a decay is present. It appears from the plot in Figure 3.11 that such a decay is more likely than the Normal or stable Paretian hypothesis. It would seem, therefore, that the idea of introducing the exponential power distributions to allow for more flexibility in the distribution of the shocks is worthy of consideration, although it has to be kept in mind that the other explanations we mentioned earlier in section 3.2 could account for this phenomenon.

The joint posterior distribution $p(\beta,\theta|\underset{\sim}{z})$ is derived and its contours corresponding to highest posterior density regions are plotted in Figure 3.12. The conditional posterior distributions $p(\theta|\underset{\sim}{z},\beta)$ are plotted for various β in Figure 3.13 and they show moderate sensitivity to changes in β. Assuming a uniform reference prior for β, the posterior distribution $p(\beta|\underset{\sim}{z})$ is derived and is given in Figure 3.14.

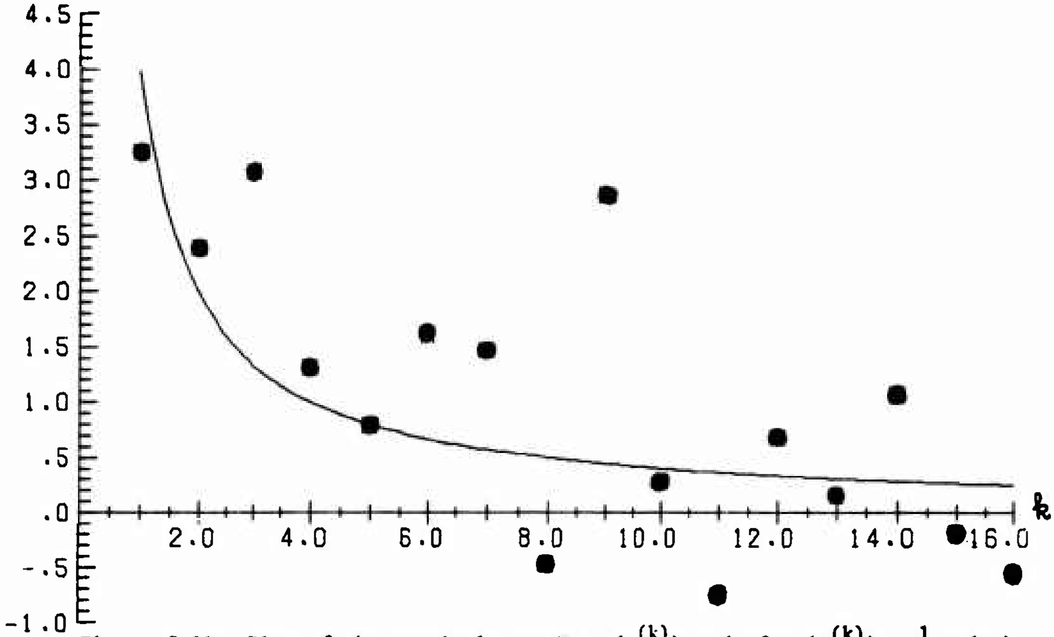


Figure 3.11: Plot of the sample kurtosis $r_2(w_t^{(k)})$ and of $\gamma_2(w_t^{(k)}) = \frac{1}{k}\gamma_2(w_t)$ with $\gamma_2(w_t) = 4$: Series B.

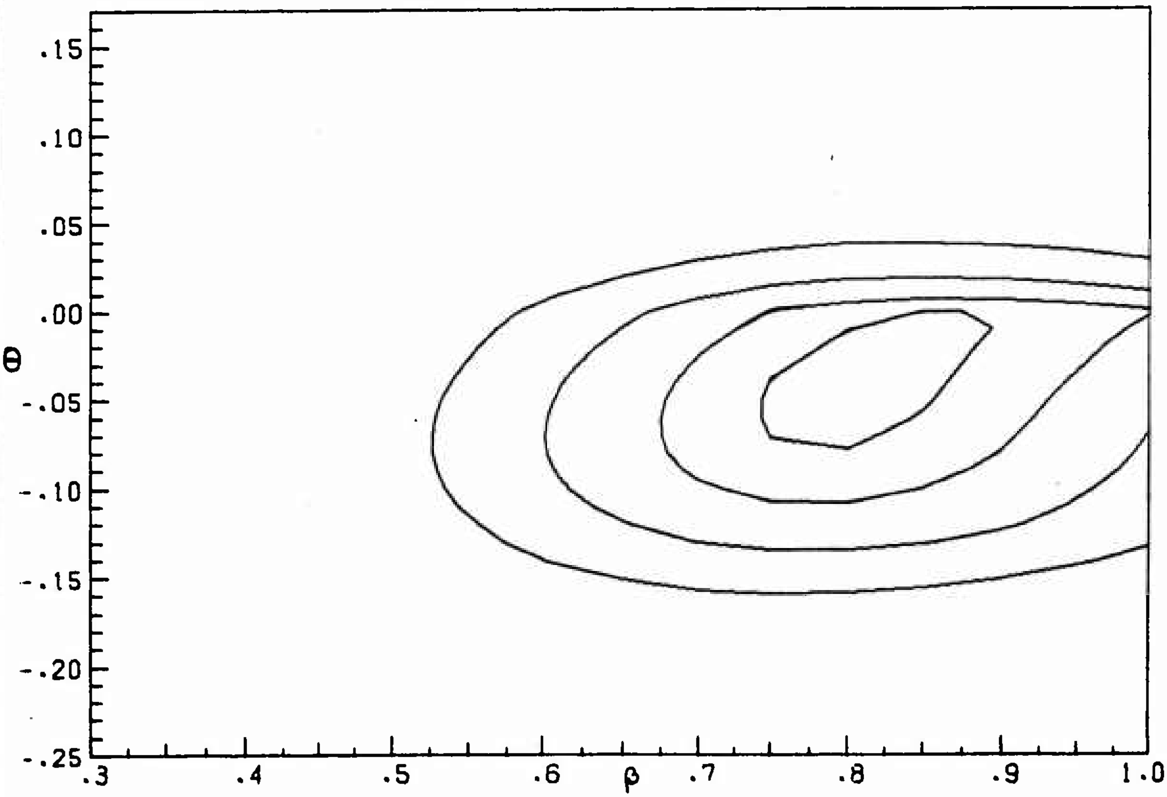


Figure 3.12: Contours of the posterior distribution of (β,θ) corresponding roughly to

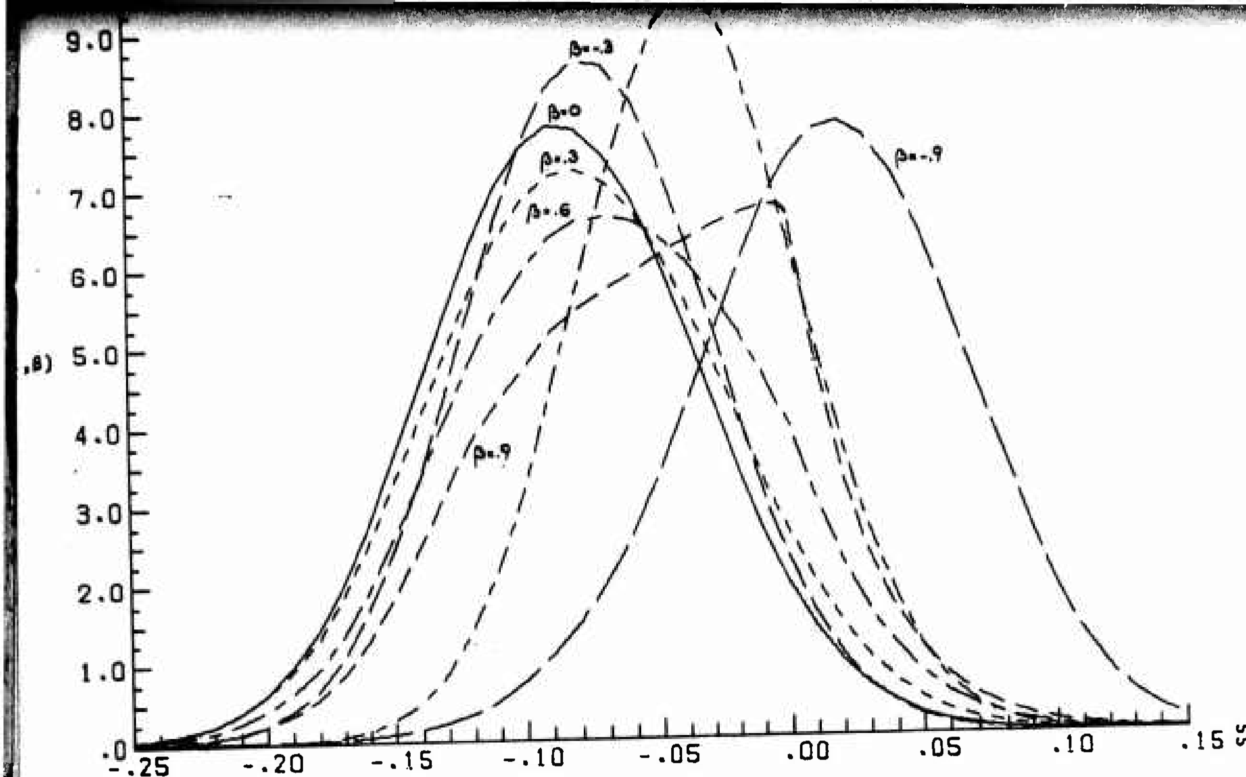


Figure 3.13: Posterior distribution of θ for various fixed β: Series B.

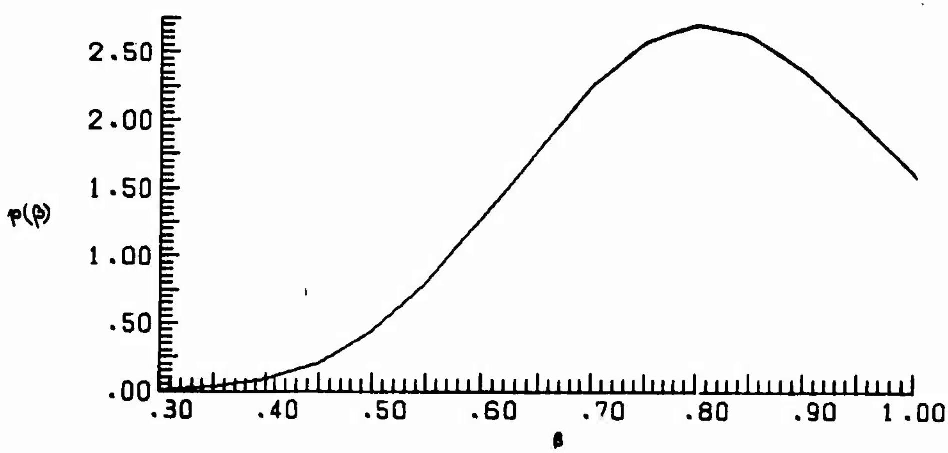


Figure 3.14: Posterior distribution of β assuming a uniform reference prior: Series B.

e posterior distribution $p(\theta|\underset{\sim}{z})$ is compared to the osterior distribution of the moving average parameter ssuming Normal distributed errors and both are plotted in gure 3.15.

Again, as in series A, the forecasts $\hat{z}_n(\ell)$ of series B e very insensitive to the choice of β. This can be seen om the predictive distributions $p(z_{n+1}|\underset{\sim}{z},\beta)$ which are ven in Figure 3.16 for various β. As before these have milar means, but differ considerably with respect to e shape of the predictive distribution. In Figure 3.17 compare the predictive distribution $p(z_{n+1}|\underset{\sim}{z})$ with $(z_{n+1}|\underset{\sim}{z},\beta=0)$, the predictive distribution assuming Normal rors. The difference in the shape of the predictive stributions given in Figure 3.17 is very marked. To get e idea about the non Normality of forecasts more step- ead, we derive the kurtosis of the ℓ-step ahead forecast ror $\gamma_2(e_n(\ell))$ using the mode $(\hat{\beta},\hat{\theta}) = (-.04,.80)$ as timates for β and θ. This is given in Figure 3.18.

8 Concluding remarks:

In section 3.6 we saw that for the integrated first der moving average process

$$p(z_{n+\ell}|\underset{\sim}{z},\beta) = \int_{-1}^{1} p(z_{n+\ell}|\underset{\sim}{z},\beta,\theta)\, p(\theta|\underset{\sim}{z},\beta)d\theta \qquad (3.8.1)$$

e predictive distribution for given β and θ is symmetric ound $\hat{z}_n(\ell)$, and the shape of the distribution depends on

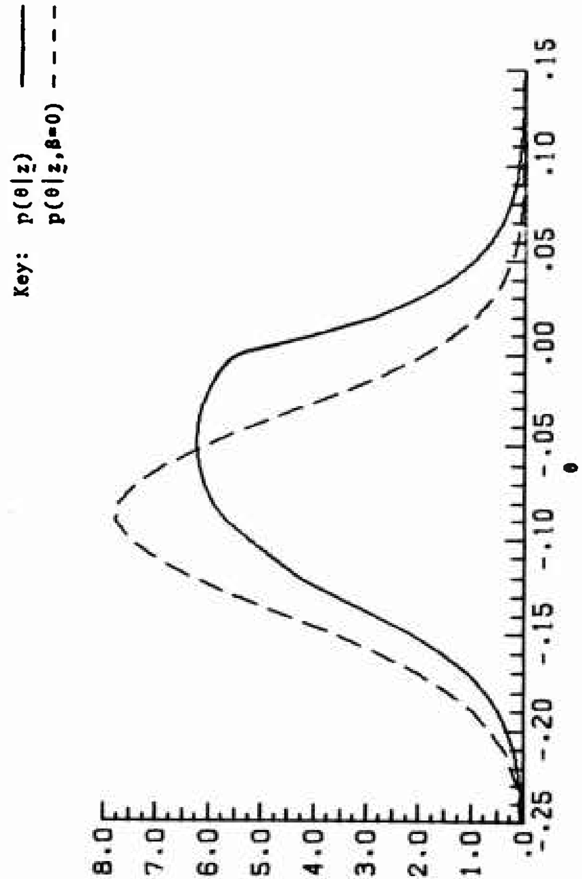


Figure 3.15: Posterior distribution of θ, $p(\theta|\underset{\sim}{z})$, and posterior distribution of θ assuming Normal distributed shocks, $p(\theta|\underset{\sim}{z},\beta=0)$: Series B.

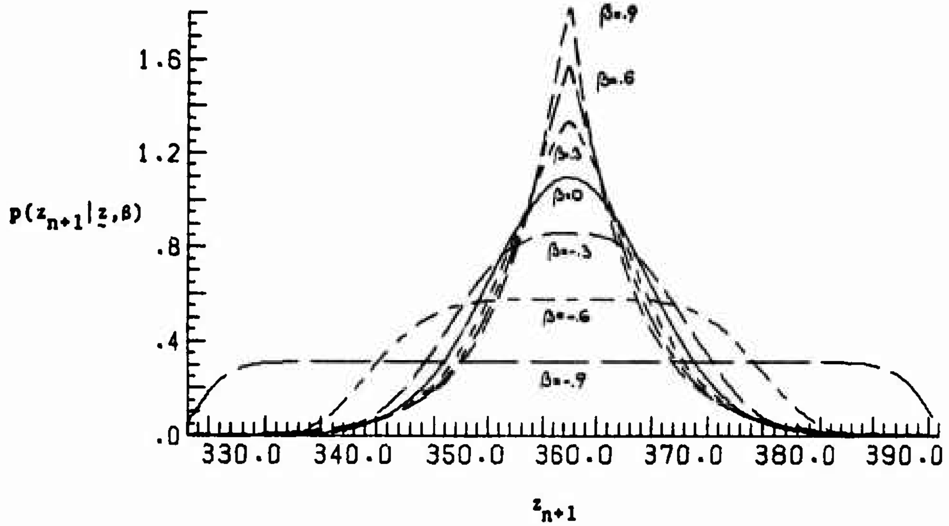


Figure 3.16: Posterior one step ahead predictive distribution for various fixed β: Series B, n=369.

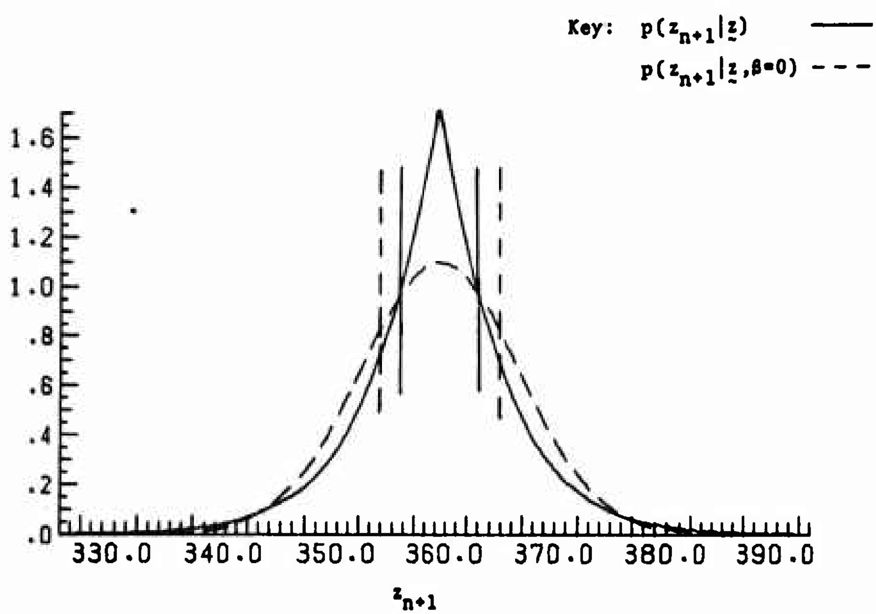


Figure 3.17: Posterior one step ahead predictive distribution $p(z_{n+1}|\underset{\sim}{z})$ compared with $p(z_{n+1}|\underset{\sim}{z},\beta=0)$, the posterior one step ahead predictive distribution assuming Normal distributed shocks, and their 50% probability limits: Series B

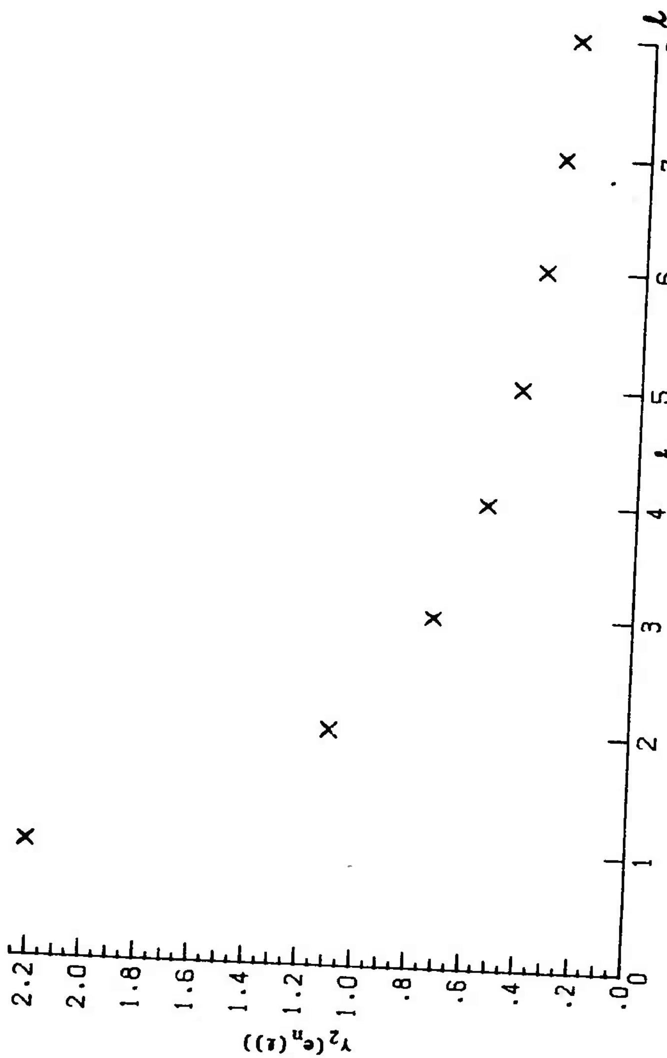


Figure 3.18: Plot of the kurtosis of the ℓ-step ahead forecast error: Series B.

the value of β.

For the two examples in section 3.7 the inference about the parameter θ changes for different assumptions about the distribution of the shocks a_t. The forecasts $\hat{z}_n(\ell)$, however, do not change much over the range of θ over which the conditional posterior distributions $p(\theta|\underset{\sim}{z},\beta)$ are centered. In this case, even though there are differences in $p(\theta|\underset{\sim}{z},\beta)$ for different values of β, the modes of the predictive distributions in (3.8.1) are similar and thus insensitive to changes in β.

The shape of the predictive distribution depends on the value of β. For the one step ahead predictive distribution we compare $p(z_{n+1}|\underset{\sim}{z})$ and $p(z_{n+1}|\underset{\sim}{z},\beta=0)$ to show the effect of non Normality of the shocks on the shape of the distribution. Furthermore the kurtosis of the forecast errors more than one step ahead is computed and plotted. Since both considered series are non stationary, the ψ-weights do not die out, and it is seen that even for moderate lead times the kurtosis becomes small fairly quickly.

APPENDIX 3.1

Nonstationary models leading to leptokurtic error distributions.

Model 1: $w_t = (1-\theta B)a_t \quad Ea_t = 0$ for all t

$$Ea_t^2 \begin{cases} \sigma_1^2 & \text{for } 1\le t\le T_1 \\ \sigma_2^2 & \text{for } T_1+1\le t\le T_1+T_2 \end{cases}$$

Model 2: $w_t = (1-\theta_t B)a_t \quad Ea_t = 0$ and $Ea_t^2 = \sigma_a^2$ for all t

$$\theta_t = \begin{cases} \theta_1 & \text{for } 1\le t\le T_1 \\ \theta_2 & \text{for } T_1+1\le t\le T_1+T_2 \end{cases}$$

Model 3: $w_t = (1-\theta_t B)a_t \quad Ea_t = E\alpha_t = 0$

$\phi(B)(\theta_t-\theta) = \psi(B)\alpha_t \qquad Ea_t^2 = \sigma_a^2;\ E\alpha_t^2 = \sigma_\alpha^2;$ for all t

$\{a_t\}$ and $\{\alpha_t\}$ are independent white noise sequences.

Then it can be shown that in everyone of these cases the errors e_t, which are evaluated from the model with constant parameters $w_t = (1-\theta B)e_t$, will have non negative kurtosis, thus leading to leptokurtic distributions of the shocks. Furthermore, in order to be able to ignore end effects, we assume that T_1 and T_2 are large and $\frac{T_1}{T_1+T_2} = k$.

i) For model 1: $e_t = a_t$ with $Ea_t^2 = \begin{cases} \sigma_1^2 & 1\le t\le T_1 \\ \sigma_2^2 & T_1+1 \le t\le T_1+T_2 \end{cases}$

$$E\left[\frac{1}{T_1+T_2}\sum_1^{T_1+T_2} e_t^2\right] = k\sigma_1^2 + (1-k)\sigma_2^2 \qquad \text{(A.3.1}$$

$$E\left[\frac{1}{T_1+T_2}\sum_1^{T_1+T_2} e_t^4\right] = 3[k\sigma_1^4 + (1-k)\sigma_2^4] \qquad \text{(A.3.1}$$

and

$$\gamma_2(e_t) = \frac{3[k\sigma_1^4+(1-k)\sigma_2^4]}{[k\sigma_1^2+(1-k)\sigma_2^2]^2} - 3 = \frac{3k(1-k)(\sigma_1^2-\sigma_2^2)^2}{[k\sigma_1^2+(1-k)\sigma_2^2]^2} \ge (\qquad \text{(A.3.}$$

ii) For model 2: $(1-\theta B)e_t = (1-\theta_t b)a_t$ and $Ea_t^2 = \sigma_a^2$ for all t; $\theta_t = \begin{cases} \theta_1 & 1\le t\le T_1 \\ \theta_2 & T_1+1\le t\le T_1+T_2 \end{cases}$

We expand e_{T_1+n} for $n > 0$ as a function of previous shock: We also note that the regime with parameter θ_1 provides t starting values for the regime with parameter θ_2.

$$e_{T_1+n} = a_{T_1+n}+(\theta-\theta_2)a_{T_1+n-1}+\dots+ \theta^{n-2}(\theta-\theta_2)a_{T_1+1}-\theta^{n-1}\theta_2 a_{T_1}+\theta^n e_{T_1} \qquad \text{(A.3}$$

Furthermore $e_{T_1} = [1+\sum_{j=0}^{\infty}\theta^j(\theta-\theta_1)b^{j+1}]a_{T_1}$ and substitutin

into (A.3.1.4) yields

$$e_{T_1+n} = a_{T_1+n}+(\theta-\theta_2)a_{T_1+n-1}+\cdots$$

$$+\theta^{n-1}(\theta-\theta_2)a_{T_1}+\theta^{n}[(\theta-\theta_1)\sum_{j=0}^{\infty}\theta^{j}a_{T_1-1-j}] \quad \text{(A.3.1.5)}$$

$$E(e^2_{T_1+n}) = \sigma_a^2[1+(\theta-\theta_2)^2\,\frac{1-\theta^{2n}}{1-\theta^2}+\theta^{2n}(\theta-\theta_1)^2\frac{1}{1-\theta^2}]$$

$$= \sigma_a^2\{1+\frac{(\theta-\theta_2)^2}{1-\theta^2}+\frac{\theta^{2n}}{1-\theta^2}[(\theta-\theta_1)^2-(\theta-\theta_2)^2]\}$$

and

$$E[\frac{1}{T_1+T_2}\sum_{1}^{T_1+T_2} e_t^2] = \frac{1}{T_1+T_2}[\sum_{n=1}^{T_2}E(e^2_{T_1+n})+\sum_{t=1}^{T_1}E(e_t^2)]$$

$$= \frac{1}{T_1+T_2}\{[1+\frac{(\theta-\theta_2)^2}{1-\theta^2}]T_2+\frac{\theta^2(1-\theta^{2T_2})}{(1-\theta^2)^2}$$

$$[(\theta-\theta_1)^2-(\theta-\theta_2)^2]+[1+(\theta-\theta_1)^2\frac{1}{1-\theta^2}]T_1\}$$

$$E(\frac{1}{T_1+T_2}\sum_{1}^{T_1+T_2} e_t^2) = k(1+\frac{(\theta-\theta_1)^2}{1-\theta^2})+(1-k)(1+\frac{(\theta-\theta_2)^2}{1-\theta^2})$$

$$+ 0(\frac{1}{T_1+T_2}) \quad \text{(A.3.1.6)}$$

It can be shown that:

$$E[\frac{1}{T_1+T_2}\sum_{n=1}^{T_2} e^4_{T_1+n}] = \frac{1}{T_1+T_2}\{3\sigma_a^4[T_2+\frac{(\theta-\theta_2)^4}{1-\theta^4}T_2]$$

$$+6\sigma_a^4[\frac{(\theta-\theta_2)^2}{1-\theta^2}T_2+\frac{(\theta-\theta_2)^4\theta^2}{(1-\theta^2)(1-\theta^4)}T_2]\}+0(\frac{1}{T_1+T_2})$$

$$= \frac{3T_2\sigma_a^4}{T_1+T_2}\{1+\frac{(\theta-\theta_2)^4}{1-\theta^4}[1+\frac{2\theta^2}{1-\theta^2}]+2\frac{(\theta-\theta_2)^2}{1-\theta^2}\}+0(\frac{1}{T_1+T_2})$$

$$= 3\sigma_a^4(1-k)\{1+\frac{(\theta-\theta_2)^2}{1-\theta^2}\}^2 + 0(\frac{1}{T_1+T_2}) \quad \text{(A.3.1.7)}$$

Similarily, it can be shown that

$$E[\frac{1}{T_1+T_2}\sum_{t=1}^{T_1} e_t^4] = 3\sigma_a^4k\{1+\frac{(\theta-\theta_1)^2}{1-\theta^2}\}^2 \quad \text{(A.3.1.8)}$$

Ignoring end effects (T_1,T_2 large),

$$E(\frac{1}{T_1+T_2}\sum_{1}^{T_1+T_2} e_t^4) = 3\sigma_a^4\{k[1+\frac{(\theta-\theta_1)^2}{1-\theta^2}]^2$$

$$+ (1-k)\;[1+\frac{(\theta-\theta_2)^2}{1-\theta^2}]^2\} \quad \text{(A.3.1.9)}$$

and

$$\gamma_2(z_t) = \frac{3\sigma_a^4\{k[1+\frac{(\theta-\theta_1)^2}{1-\theta^2}]^2+(1-k)[1+\frac{(\theta-\theta_2)^2}{1-\theta^2}]^2\}}{\sigma_a^4\{k[1+\frac{(\theta-\theta_1)^2}{1-\theta^2}]+(1-k)[1+\frac{(\theta-\theta_2)^2}{1-\theta^2}]\}^2} - 3 =$$

$$= \frac{3k(1-k)\{[1+\frac{(\theta-\theta_1)^2}{1-\theta^2}]^2-[1+\frac{(\theta-\theta_2)^2}{1-\theta^2}]^2\}^2}{\{k[1+\frac{(\theta-\theta_1)^2}{1-\theta^2}]+(1-k)[1+\frac{(\theta-\theta_2)^2}{1-\theta^2}]\}^2} \geq 0$$

(A.3.1.10

iii) For model 3: We consider the case where

$$\phi(B) = \psi(B) = 1; \text{ i.e.: } \theta_t = \theta + \alpha_t$$

$$(1-\theta B)e_t = w_t = (1-\theta_t B)a_t = (1-\theta B)a_t - \alpha_t a_{t-1}$$

$$e_t = a_t - \frac{\alpha_t a_{t-1}}{1-\theta B}$$

$$E(e_t^2) = \sigma_a^2\{1+\sigma_\alpha^2 \frac{1}{1-\theta^2}\} \qquad (A.3.1.11)$$

$$E(e_t^4) = 3\sigma_a^4\{1+2\sigma_\alpha^2 \frac{1}{1-\theta^2} + \sigma_\alpha^4 \frac{3-\theta^2}{(1-\theta^4)(1-\theta^2)} \qquad (A.3.1.12)$$

and

$$\gamma_2(e_t) = \frac{6\sigma_\alpha^4}{(1-\theta^4)(1+\sigma_\alpha^2 \frac{1}{1-\theta^2})^2} \geq 0 \qquad (A.3.1.13)$$

Q.E.D.

APPENDIX 3.2

Moments and characteristic function of the exponential power distribution centered around zero

$$p(x) = \omega(\beta)\sigma^{-1} \exp\{-\frac{c(\beta)}{\sigma^{2/1+\beta}} |x|^{2/1+\beta}\}$$

$$-\infty < x < \infty; \quad -1 < \beta \leq 1; \ \sigma > 0$$

$$\omega(\beta) = \frac{\{\Gamma(\frac{3}{2}(1+\beta))\}^{1/2}}{(1+\beta)\{\Gamma(\frac{1}{2}(1+\beta))\}^{3/2}} \qquad (A.3.2.1)$$

$$c(\beta) = \left\{\frac{\Gamma(\frac{3}{2}(1+\beta))}{\Gamma(\frac{1}{2}(1+\beta))}\right\}^{\frac{1}{1+\beta}}$$

We show that:

$$E(x^m) = \begin{cases} 0 & \text{if m odd integer} \\ \sigma^m \dfrac{\{\Gamma(\frac{1}{2}(1+\beta))\}^{\frac{m}{2}-1}}{\{\Gamma(\frac{3}{2}(1+\beta))\}^{\frac{m}{2}}} \Gamma(\frac{m+1}{2}(1+\beta)) & \end{cases} \qquad (A.3.2.2)$$

$$\phi(u) = \int_{-\infty}^{+\infty} e^{iux}\, p(x)dx = \{\Gamma(\tfrac{1}{2}(1+\beta))\}^{-1}$$

$$\sum_{j=0}^{\infty} \frac{\left\{\dfrac{iu\sigma[\Gamma(\frac{1}{2}(1+\beta))]^{1/2}}{[\Gamma(\frac{3}{2}(1+\beta))]^{1/2}}\right\}^{2j}}{(2j)!} \Gamma\{(j+\tfrac{1}{2})(1+\beta)\} \qquad (A.$$

Proof: Since the distribution in (A.3.2.1) is symmetric around zero, the odd moments vanish. If m is even

$$E(x^m) = 2\omega(\beta)\sigma^{-1}\int_0^{\infty} x^m \exp[-\frac{c(\beta)}{\sigma^{2/1+\beta}} x^{\frac{2}{1+\beta}}]dx =$$

$$= \sigma^m\, \omega(\beta)(1+\beta)\Gamma(\tfrac{m+1}{2}(1+\beta))c(\beta)^{-\frac{m+1}{2}(1+\beta)} =$$

$$= \sigma^m \frac{\{\Gamma(\frac{3}{2}(1+\beta))\}^{\frac{1}{2}}}{(1+\beta)\{\Gamma(\frac{1}{2}(1+\beta))\}^{\frac{3}{2}}}(1+\beta)\Gamma(\tfrac{m+1}{2}(1+\beta))$$

$$\times \frac{\{\Gamma(\frac{1}{2}(1+\beta))\}^{\frac{m+1}{2}}}{\{\Gamma(\frac{3}{2}(1+\beta))\}^{\frac{m+1}{2}}}$$

$$= \sigma^m \frac{\{\Gamma(\frac{1}{2}(1+\beta))\}^{\frac{m}{2}-1}}{\{\Gamma(\frac{3}{2}(1+\beta))\}^{\frac{m}{2}}}\Gamma(\tfrac{m+1}{2}(1+\beta)) \quad \text{Q.E.D.}$$

Furthermore:

$$\phi(u) = \omega(\beta)\sigma^{-1}\int_{-\infty}^{+\infty} e^{iux} \exp[-\frac{c(\beta)}{\sigma^{2/1+\beta}} |x|^{2/1+\beta}]dx =$$

$$= \{\Gamma(\tfrac{1}{2}(1+\beta))\}^{-1}\sum_{j=0}^{\infty} \frac{(iu)^{2j}}{(2j)!}\sigma^{2j}\left\{\frac{\Gamma(\frac{1}{2}(1+\beta))}{\Gamma(\frac{3}{2}(1+\beta))}\right\}^j$$

$$\times \Gamma(\tfrac{2j+1}{2}(1+\beta))$$

$$= \Gamma\{\tfrac{1}{2}(1+\beta)\}^{-1}\sum_{j=0}^{\infty}\frac{\left\{\frac{iu\sigma[\Gamma(\frac{1}{2}(1+\beta))]^{1/2}}{[\Gamma(\frac{3}{2}(1+\beta))]^{1/2}}\right\}^{2j}}{(2j)!}$$

$$\times \Gamma((j+\tfrac{1}{2})(1+\beta)) \qquad \text{Q.E.D.}$$

If $\beta = 0$ (Normal distribution)

$$\phi(u) = \{\Gamma(\tfrac{1}{2})\}^{-1}\sum_{j=0}^{\infty}\frac{(iu\sigma\sqrt{2})^{2j}}{(2j)!}\Gamma(j+\tfrac{1}{2}).$$

Since $\dfrac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} = \dfrac{(2j)!}{2^{2j}}$ we get $\phi(u) = \displaystyle\sum_{j=0}^{\infty}\frac{(\frac{-u^2\sigma^2}{2})^j}{j!} = e^{-\frac{u^2\sigma^2}{2}}$

(A.3.2.4)

If $\beta = 1$ (double exponential distribution)

$$\phi(u) = \sum_{j=0}^{\infty}\frac{(\frac{iu\sigma}{\sqrt{2}})^{2j}}{(2j)!}\Gamma(2j+1) = \sum_{j=0}^{\infty}(\frac{i^2u^2\sigma^2}{2})^j = \frac{1}{1+\frac{u^2\sigma^2}{2}}$$

(A.3.2.5)

The Normal distribution is invariant under addition, however in general this invariance property does not hold for the family of the exponential power distributions. Considering the double exponential distribution ($\beta=1$) we see that

$$p(x_i) = \tfrac{1}{2}e^{-\sqrt{2}|x_i|} \quad \text{for } i = 1,2 \quad (\sigma=1)$$

and the distribution of $z = x_1+x_2$ is given by

$$p(z) = \frac{1}{2\sqrt{2}} e^{-\sqrt{2}|z|} + \frac{1}{2}|z| e^{-\sqrt{2}z}$$

which is not a member of the family of exponential power distributions.

APPENDIX 3.3

Noninformative prior distribution for an AR(p) process from the family of symmetric exponential power distributions.

In this appendix we present an argument for choosing a particular metric in terms of which a locally uniform prior can be regarded as noninformative about the parameters in (3.4.4). We apply Jeffreys' rule, which says that the prior distribution for a set of parameters is approximately noninformative if it is taken proportional to the square root of the determinant of Fisher's information matrix.

For the class of exponential power distribution the information matrix exists only for $\beta < 0$. Jeffreys' rule, like most rules, should not be applied mechanically. The result of this appendix, for the case when $\beta < 0$, will however serve as a guideline for choosing the prior.

The log likelihood for the AR(p) process follows from (3.4.4) and is given by

$$\ell(\sigma,\underline{\phi}) = -(n-p)\log\sigma - \frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=p+1}^{n} |z_t - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p}|^{\frac{2}{1+\beta}} \tag{A.3.3.1}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n-p}{\sigma} + \frac{2}{1+\beta} \frac{c(\beta)}{\sigma^{\frac{2}{1+\beta}+1}} \sum_{t=p+1}^{n} |z_t - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p}|^{\frac{2}{1+\beta}}$$

$$\frac{\partial^2 \ell}{\partial\sigma^2} = \frac{n-p}{\sigma^2} - \frac{2}{1+\beta}\left[\frac{2}{1+\beta}+1\right] \frac{c(\beta)}{\sigma^{\frac{2}{1+\beta}+2}}$$

$$\sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}}$$

Thus

$$E\{-\frac{\partial^2 \ell}{\partial\sigma^2}\} = -\frac{n-p}{\sigma^2} + \frac{2}{1+\beta}\left[\frac{2}{1+\beta}+1\right] \frac{c(\beta)}{\sigma^{\frac{2}{1+\beta}+2}} (n-p)\, E|a|^{\frac{2}{1+\beta}} \tag{A.3.3.2}$$

Using the result in Appendix 3.2,

$$E|a|^{\frac{2}{1+\beta}} = \sigma^{\frac{2}{1+\beta}} \frac{\{\Gamma(\frac{1}{2}(1+\beta))\}^{\frac{1}{1+\beta}-1}\,\Gamma(\frac{1}{2}(3+\beta))}{\{\Gamma(\frac{3}{2}(1+\beta))\}^{\frac{1}{1+\beta}}}$$

Therefore:

$$E(-\frac{\partial^2 \ell}{\partial\sigma^2}) = \frac{n-p}{\sigma^2}\left[\frac{\Gamma(\frac{1}{2}(3+\beta))}{\Gamma(\frac{1}{2}(1+\beta))}\,\frac{2(3+\beta)}{(1+\beta)^2} - 1\right] \tag{A.3.3.3}$$

Furthermore it is easily shown that

$$\frac{\partial^2 \ell}{\partial\sigma\partial\phi_i} = \left(\frac{2}{1+\beta}\right)^2 \frac{c(\beta)}{\sigma^{2/1+\beta}}$$

$$\sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}-1} z_{t-i}\delta_t$$

where $\delta_t = \begin{cases} -1 & \text{if } z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p} > 0 \\ +1 & \text{if } z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p} < 0 \end{cases}$

and therefore

$$E\{-\frac{\partial^2 \ell}{\partial\sigma\partial\phi_i}\} = 0 \qquad 1 \le i \le p \tag{A.3.3.4}$$

Similarly,

$$\frac{\partial^2 \ell}{\partial\phi_i\partial\phi_j} = -\frac{2}{1+\beta}\left[\frac{2}{1+\beta}-1\right] \frac{c(\beta)}{\sigma^{\frac{2}{1+\beta}}}$$

$$\sum_{t=p+1}^{n} |z_t-\phi_1 z_{t-1}-\cdots-\phi_p z_{t-p}|^{\frac{2}{1+\beta}-2} z_{t-i}z_{t-j}$$

and

$$E\{-\frac{\partial^2 \ell}{\partial\phi_i\partial\phi_j}\} = \frac{2(1-\beta)}{(1+\beta)^2}\,\frac{c(\beta)}{\sigma^{\frac{2}{1+\beta}}}(n-p)\; E|a|^{\frac{2}{1+\beta}-2}\,\gamma_{|i-j|} \tag{A.3.3.5}$$

where $\gamma_{|i-j|} = E(z_{t-i}z_{t-j})$ and $\rho_{|i-j|} = \frac{\gamma_{|i-j|}}{\gamma_0}$

Substituting

$$E|a|^{\frac{2}{1+\beta}-2} = \sigma^{\frac{2}{1+\beta}-2} \frac{\{\Gamma(\frac{3}{2}(1+\beta))\}^{\frac{\beta}{1+\beta}}}{\{\Gamma(\frac{1}{2}(1+\beta))\}^{\frac{\beta}{1+\beta}-1}}\Gamma(\tfrac{1}{2}(1-\beta))$$

into (A.3.3.5) yields

$$E\{-\frac{\partial^2 \ell}{\partial\phi_i\partial\phi_j}\} = (n-p)\,\frac{2(1-\beta)}{(1+\beta)^2}\,\frac{\Gamma(\frac{3}{2}(1+\beta))}{\{\Gamma(\frac{1}{2}(1+\beta))\}^2}\Gamma(\tfrac{1}{2}(1-\beta))\rho_{|i-j|} \tag{A.3.3.6}$$

Therefore, the determinant of Fisher's information matrix is

$$|I(\sigma,\underset{\sim}{\phi})| = I(\sigma)|I(\underset{\sim}{\phi})| = [\frac{\Gamma(\frac{1}{2}(3+\beta))\ 2(3+\beta)}{\Gamma(\frac{1}{2}(1+\beta))\ (1+\beta)^2} - 1]$$

$$[\frac{2(1-\beta)}{(1+\beta)^2} \frac{\Gamma(\frac{3}{2}(1+\beta))}{\{\Gamma(\frac{1}{2}(1+\beta))\}^2} \Gamma(\frac{1}{2}(1-\beta))]^p \sigma^{-2} |P_p|$$

where $\underset{\sim}{\phi}' = (\phi_1,\ldots,\phi_p)$ and P_p is a p×p correlation matrix with elements $\rho_{|i-j|}$ $1 \leq i, j \leq p$. For given $\beta < 0$ the noninformative Jeffreys' prior distribution is therefore given by

$$p(\sigma,\underset{\sim}{\phi}) = \sigma^{-1}|P_p|^{\frac{1}{2}} \qquad (A.3.3.7)$$

REFERENCES

Bachelier, L. (1900), "Theory of Specutlation," reprinted in Cootner, P. H. (ed.) [1964], The Random Character of Stock Market Prices," MIT Press, Massachusetts.

Bloomfield, P. (1972), "On the error of prediction of a time series," Biometrika, 59, 501-507.

Box, G. E. P. and Jenkins, G. M. (1970), Time Series Analysis, Forecasting and Control, Holden-Day; San Francisco.

Box, G. E. P. (1953) "A note on regions for test of kurtosis," Biometrika, 40, 465-469.

Box, G. E. P. and Cox, D. R. (1964), "An analysis of transformations," Jour. of Royal Statis. Soc., Series B, 26, 211-252.

Box, G. E. P. and Tiao, G. C. (1962), "A further look at robustness via Bayes theorem," Biometrika, 49, 419-433.

Box, G. E. P. and Tiao, G. C. (1973), Bayesian Inference in Statistical Analysis, Addison-Wesley; Reading, Massachusetts.

Fama, E. F. (1965), "The behavior of stock market prices," Journal of Business, 38, 34-105.

Fama, E. F. and Roll, R. (1968), "Some properties of symmetric stable distributions," JASA, 63, 817-836.

Hsu, Der-Ann (1973), "Stochastic instability and the behavior of stock prices," Ph.D. thesis, University of Wisconsin-Madison.

Hsu, Der-Ann, Miller, R. B. and Wichern, D. W. (1974), "On the Paretian stable behavior of stock market prices," JASA, 69, 108-113.

Kendall, M. G. (1953), "The analysis of economic time series, Part I: Prices," Jour. of Royal Statis. Soc., Series A, 116, 11-34.

Lund, D. R. (1967), "Parameter estimation in a class of power distributions," Ph.D. thesis, University of Wisconsin-Madison.

Mandelbrot, B. (1961), "Stable Paretian random functions and the multiplicative variation of income," Econometrica, 29, 517-543.

Mandelbrot, B. (1963), "The variation of certain speculative prices," Jour. of Business, 36, 394-419.

Mandelbrot, B. (1967), "The variation of some other speculative prices," Jour. of Business, 40, 393-413.

Mandelbrot, B. and Taylor, H. M. (1967), "On the distribution of stock price difference," Operations Res., 15, 1057-1062.

Miller, R. B., Wichern, D. W. and Hsu, Der-Ann (1971), "Changes of variance in time series models," Technical Report #284, Department of Statistics, University of Wisconsin-Madison.

Praetz, P. D. (1972), "The distribution of share price changes," Jour. of Business, 45, 49-55.

Press, S. J. (1967), "A compound events model for security prices," Jour. of Business 40, 317-335.

Samuelson, P. A. (1971), "Mathematics of speculative price," Fall Meeting of the Society for Industrial and Applied Mathematics, Madison, Wis., Oct. 11-13, 1971.

Wichern, D. W. (1969), "Studies in the identification and forecasting of non stationary time series," Ph.D. thesis, University of Wisconsin-Madison.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR - TR - 76 - 0725	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TOPICS IN TIME SERIES ANALYSIS III. ARIMA time series models with non Normal shocks		5. TYPE OF REPORT & PERIOD COVERED Scientific Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Johannes Ledolter and George E. P. Box		8. CONTRACT OR GRANT NUMBER(s) AFOSR-AF-72-2363-C
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics 1210 W. Dayton Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Building 410 Bolling AFB, D. C., 20332		12. REPORT DATE Jan., 1976
		13. NUMBER OF PAGES 77
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ARIMA models
stock price series
non Normal shocks
exponential power distributions
kurtosis
predictive distribution
robustness

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper the parent distribution of the shocks a_t in ARIMA models is extended to the family of symmetric exponential power distributions, which covers the Normal as well as particular forms of symmetric leptokurtic and platykurtic distributions. The inference robustness of the parameters of ARIMA models and the effect of non Normality of the shocks on the forecasts is studied.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR - TR - 76 - 0725
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): 6 TOPICS IN TIME SERIES ANALYSIS III. ARIMA TIME SERIES MODELS WITH NON-NORMAL SHOCKS,

5. TYPE OF REPORT & PERIOD COVERED: 9 Interim rept.,

6. PERFORMING ORG. REPORT NUMBER: ~~Tech Report No 448~~

7. AUTHOR(s): 10 Johannes Ledolter [redacted] George E. P. Box

8. CONTRACT OR GRANT NUMBER(s): 15 ~~AF~~-AFOSR [redacted]-2363-72

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
University of Wisconsin
Department of Statistics
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 9769-05

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research/NM
Bolling AFB, Washington, DC 20332

12. REPORT DATE: 11 Jan [redacted] 76

13. NUMBER OF PAGES: 77

12 45p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 14 UWIS-DS-76-448

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

16 AF-9769 17 976905

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 AFOSR 19 TR-76-0725

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ARIMA models	kurtosis
stock price series	predictive distribution
non Normal shocks	robustness
exponential power distributions	

a sub t.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper the parent distribution of the shocks a_t in ARIMA models is extended to the family of symmetric exponential power distributions, which covers the Normal as well as particular forms of symmetric leptokurtic and playkurtic distributions. The inference robustness of the parameters of ARIMA models and the effect of non Normality of the shocks on the forecasts is studied.